

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/89288>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

**A corpus-driven investigation into the semantic
patterning of grammatical keywords in
undergraduate History and PIR (Politics &
international Relations) essays**

Karin Whiteside

A thesis submitted to the University of Warwick for the degree of
Doctor of Philosophy in Applied Linguistics

**Centre for Applied Linguistics
University of Warwick
July 2016**

Contents

| | |
|---|------|
| Figures & Tables | v |
| Acknowledgements | xi |
| Declaration | xii |
| Abstract | xiii |
| Chapter 1: Introduction | 1 |
| Chapter 2: Literature Review | 7 |
| 2.1. Genre | 8 |
| 2.1.1. New Rhetoric and genre | 9 |
| 2.1.2. SFL and genre | 10 |
| 2.1.3. ESP and genre | 13 |
| 2.1.4. Paltridge and Bruce | 16 |
| 2.1.5. Changing views on the ‘three schools’ | 19 |
| 2.2. Discipline | 21 |
| 2.2.1. A model for the disciplines | 21 |
| 2.2.2. Discourse community and disciplinary discourse | 25 |
| 2.2.3. Disciplinary variation in professional academic writing | 28 |
| 2.2.4. Student writing, community and discipline | 37 |
| 2.2.5. Research into disciplinary variation in undergraduate student writing ... | 42 |
| 2.3. Corpus linguistics and disciplinary discourse analysis | 53 |
| 2.3.1. The nature of corpus linguistics | 54 |
| 2.3.2. The ‘corpus-based’ versus ‘corpus-driven’ debate | 57 |
| 2.3.3. The usefulness of corpus linguistics to discourse analysis, particularly in ESP | 61 |
| 2.3.4. The usefulness of keywords to discourse analysis | 63 |
| 2.3.5. Choice of items for further analysis from a keyword list: an argument for focus on ‘grammatical’ items | 65 |
| Chapter 3: Methodology | 69 |
| 3.1. Approach | 69 |
| 3.2. Research questions | 69 |
| 3.3. Data selection | 70 |
| 3.4. Data collection | 73 |

| | |
|--|-----|
| 3.4.1. Data from the BAWE corpus | 73 |
| 3.4.2 Data from my own institution | 75 |
| 3.5. Creation of research corpora | 78 |
| 3.5.1 Design considerations: size, representativeness and balance | 80 |
| 3.5.2 A closer look at the corpus data - the essay questions | 82 |
| 3.6. Keyword analyses | 86 |
| 3.6.1. Choice of reference corpus: the BNC Sampler - written component..... | 86 |
| 3.6.2. Keyness analysis procedure and identification of items for analysis | 89 |
| 3.7. Approach/procedure for further analysis of identified/selected keywords... | 94 |
| 3.7.1 Groom's (2007) approach | 95 |
| 3.7.2 The approach used in this study | 101 |
| Chapter 4: Results | 105 |
| 4.1 Selection of items for further analysis | 105 |
| 4.2 Analysis | 108 |
| 4.3 Of | 110 |
| 4.3.1 Semantic categories for <i>of</i> | 110 |
| 4.3.2 Semantic sequences using <i>of</i> | 112 |
| 4.3.3 Comparison/contrast across disciplinary and institutional sub-corpora | 125 |
| 4.4 And | 129 |
| 4.4.1 Semantic categories for <i>and</i> | 129 |
| 4.4.2 'Listed phenomena' | 130 |
| 4.4.3 'Addition' | 136 |
| 4.4.4 'Consequence' | 137 |
| 4.4.5 Phraseologies with <i>between</i> | 138 |
| 4.4.6 Comparison/contrast across disciplines and institutions..... | 139 |
| 4.5 That | 141 |
| 4.5.1 Semantic categories for <i>that</i> | 141 |
| 4.5.2 'The voice of others' | 142 |
| 4.5.3 'The voice of the essay writer' | 146 |
| 4.5.4 PHENOMENON + <i>that</i> + DETAIL | 150 |
| 4.5.5 Fixed expression – <i>the fact that</i> | 150 |
| 4.5.6 Comparison/contrast across disciplines and institutions..... | 152 |

| | |
|--|-----|
| 4.6. As | 155 |
| 4.6.1 Semantic categories for <i>as</i> | 155 |
| 4.6.2 ‘Interpretation’ | 155 |
| 4.6.3 ‘Reason-Result’ | 160 |
| 4.6.4 Fixed expression – <i>such as</i> – ‘Exemplification’ | 161 |
| 4.6.5. Fixed expression – <i>as well as</i> – ‘Addition’ | 162 |
| 4.6.6 ‘Attribution’ | 162 |
| 4.6.7 ‘Function’ | 164 |
| 4.6.8 ‘Role’ | 165 |
| 4.6.9 ‘Conceptualisation’ | 166 |
| 4.6.10 ‘Similarity’ | 167 |
| 4.6.11 ‘Emphasis of extent’ | 168 |
| 4.6.12 ‘Simultaneity’ | 169 |
| 4.6.13 Comparison/contrast across disciplinary and institutional sub-corpora | 170 |
| 4.7 This..... | 172 |
| 4.7.1 Comparison with Groom’s (2007) approach | 172 |
| 4.7.2 Semantic categories for <i>this</i> | 174 |
| 4.7.3 ‘Encapsulation’ | 175 |
| 4.7.4 ‘Commentary’ | 183 |
| 4.7.5 Semi-fixed expression - time reference..... | 187 |
| 4.7.6 Semi-fixed expression – <i>in this manner/respect/sense/way</i> | 188 |
| 4.7.7 Comparison/contrast across disciplinary and institutional sub-corpora | 188 |
| 4.8 Conclusion | 191 |
| Chapter 5: Discussion..... | 193 |
| 5.1 Aims and Research Questions | 193 |
| 5.2 Similarities and variations across the four disciplinary/institutional sub- corpora | 195 |
| 5.2.1 Similarities..... | 195 |
| 5.2.2 Variations | 213 |
| 5.3 The influence of discipline vs. institution | 223 |
| 5.4 Contributions to knowledge & implications..... | 224 |

| | |
|---|-----|
| 5.4.1. Coverage: operationalizing the text | 226 |
| 5.4.2 The role of language: Language as academic purpose..... | 227 |
| 5.4.3 The nature of disciplinary language & pedagogic implications..... | 230 |
| 5.4.4 The EGAP vs. ESAP debate | 236 |
| 5.5 Reflections on the Methodology | 237 |
| 5.5.1 Strengths of the methodology..... | 238 |
| 5.5.3 Limitations of the methodology | 241 |
| Chapter 6: Conclusions..... | 245 |
| 6.1 Achievements of the thesis | 245 |
| 6.2 Limitations and suggestions for future research | 247 |
| 6.3 Final thoughts..... | 249 |
| References..... | 250 |
| Appendix 1: Essay titles..... | 263 |
| Appendix 2: Essay titles categorised according to focus | 267 |
| Appendix 3: PIR3_008_R – phraseologies highlighted | 268 |
| Appendix 4: Analysis - <i>of</i> – pooled sample | 278 |

Figures & Tables

Figures

| | |
|---|-----|
| Figure 1: Knowledge and disciplinary grouping in Becher and Trowler (2001, p. 36) | 23 |
| Figure 2: Table from Groom, 2007, p. 99 | 96 |
| Figure 3: Table from Groom, 2007, p. 100 | 97 |
| Figure 4: The first two paragraphs of PIR3_008_R marked up for of, and, that, as & this phraseologies. | 107 |
| Figure 5: Reproduction of Groom's (2007, p. 187) 'Table 5:19: Grammatical and semantic roles of <i>these</i> in HistArt and LitArt' | 173 |
| Figure 6: Model of Subject role domains (Gosden, 1993, p. 63) | 217 |

Tables

| | |
|--|-----|
| Table 1: Corpus design | 73 |
| Table 2: HIST-X and HIST-Y sub-corpora | 79 |
| Table 3: POL/IR-X and POL/IR-Y sub-corpora | 79 |
| Table 4: All four sub-corpora | 80 |
| Table 5: HIST-X | 81 |
| Table 6: HIST-Y | 81 |
| Table 7: POL/IR-X | 81 |
| Table 8: POL/IR-Y | 81 |
| Table 9: 'specific' vs. 'general' focus | 83 |
| Table 10: Exploration and evaluation of the nature of phenomena | 84 |
| Table 11: Focus on specificity & ideas/theories/issues | 85 |
| Table 12: Focus on ideas & exploration/evaluation of the nature of phenomena | 85 |
| Table 13: Proportions of types of essay focus across sub-corpora | 86 |
| Table 14: Closed-class grammatical items key in each of the four corpora | 91 |
| Table 15: Closed-class grammatical items key in only one discipline | 92 |
| Table 16: Closed-class grammatical items key at only one institution across both disciplines | 92 |
| Table 17: Closed-class grammatical items key in three corpora | 93 |
| Table 18: Closed-class grammatical items key in one corpus | 93 |
| Table 19: Closed-class grammatical items key in a random pairing of the four corpora | 94 |
| Table 20: Comparison/Groom of key features of Groom's approach (2007) and the approach taken in this study | 101 |
| Table 21: Keywords occurring more than 300 times (or 150 times in POL/IR-Y) | 105 |
| Table 22: Labelling conventions for semantic motifs and semantic sequences | 109 |
| Table 23: Semantic analysis of <i>of</i> in pooled sample (*categories used by Groom (2007)) | 111 |

| | |
|--|-----|
| Table 24: PROPERTY + <i>of</i> + PHENOMENON (1)..... | 112 |
| Table 25: PROPERTY + <i>of</i> + PHENOMENON (2)..... | 113 |
| Table 26: PROPERTY + <i>of</i> + PHENOMENON (3)..... | 113 |
| Table 27: PROCESS + <i>of</i> + OBJECT (1)..... | 114 |
| Table 28: PROCESS + <i>of</i> + OBJECT (2)..... | 114 |
| Table 29: PROCESS + <i>of</i> + ACTOR (1)..... | 115 |
| Table 30: PROCESS + <i>of</i> + ACTOR (2)..... | 116 |
| Table 31: PROCESS + <i>of</i> + ACTOR (3)..... | 116 |
| Table 32: PROCESS + <i>of</i> + ACTOR (4)..... | 117 |
| Table 33: PROPORTION OR AMOUNT + <i>of</i> + PHENOMENON (1)..... | 117 |
| Table 34: PROPORTION OR AMOUNT + <i>of</i> + PHENOMENON (2)..... | 118 |
| Table 35: PROPORTION OR AMOUNT + <i>of</i> + PHENOMENON (3)..... | 118 |
| Table 36: CONCEPTUALISATION + <i>of</i> + PHENOMENON..... | 118 |
| Table 37: QUALITY + <i>of</i> + PHENOMENON..... | 119 |
| Table 38: PHENOMENON + <i>of</i> + PARTICULAR INSTANCE OR NATURE..... | 119 |
| Table 39: The difference between PHENOMENON + <i>of</i> + PARTICULAR INSTANCE OR NATURE and PROPERTY + <i>of</i> + PHENOMENON sequences..... | 120 |
| Table 40: The difference between PHENOMENON + <i>of</i> + PARTICULAR INSTANCE OR NATURE and CONCEPTUALISATION + <i>of</i> + PHENOMENON sequences..... | 120 |
| Table 41: CONCEPTUAL LABELLING + <i>of</i> + CONCEPT..... | 121 |
| Table 42: The word perception in CONCEPTUAL LABELLING + <i>of</i> + CONCEPT and PROCESS + <i>of</i> + OBJECT sequences..... | 121 |
| Table 43: The word sense in CONCEPTUAL LABELLING + <i>of</i> + CONCEPT and PHENOMENON + <i>of</i> + PARTICULAR INSTANCE OR NATURE sequences..... | 122 |
| Table 44: MODE LABEL + <i>of</i> + PROCESS..... | 122 |
| Table 45: The word system in MODE + <i>of</i> + PROCESS and PHENOMENON + <i>of</i> + PARTICULAR INSTANCE OR NATURE sequences..... | 123 |
| Table 46: AUTHORITY + <i>of</i> + DOMAIN..... | 123 |
| Table 47: PHENOMENON + <i>of</i> + TIME..... | 123 |
| Table 48: RESULT + <i>of</i> + PHENOMENON..... | 124 |
| Table 49: TEXT + <i>of</i> + CONTENT..... | 124 |
| Table 50: CLASSIFICATION LABELLING + <i>of</i> + PHENOMENON..... | 125 |
| Table 51: The word field(s) in CLASSIFICATION LABELLING + <i>of</i> + PHENOMENON and CONCEPTUALISATION + <i>of</i> + PHENOMENON sequences..... | 125 |
| Table 52: Semantic sequences for <i>of</i> - frequency and percentage in the pooled and disciplinary/institutional sub-corpora samples..... | 126 |
| Table 53: Semantic sequences for <i>of</i> - frequency and percentage in the pooled and disciplinary/institutional sub-corpora samples – adjusted percentages..... | 127 |
| Table 54: Second 300-line disciplinary/institutional samples for PHENOMENON + <i>of</i> + TIME..... | 128 |

| | |
|--|-----|
| Table 55: Second 300-line disciplinary/institutional samples for RESULT + <i>of</i> + PHENOMENON | 129 |
| Table 56: Semantic analysis of <i>and</i> in pooled sample | 130 |
| Table 57: Phraseology boundaries for <i>and</i> | 130 |
| Table 58: PROCESS + <i>and</i> + PROCESS | 131 |
| Table 59: ABSTRACT CONCEPT + <i>and</i> + ABSTRACT CONCEPT | 132 |
| Table 60: QUALITY + <i>and</i> + QUALITY | 132 |
| Table 61: CLASSIFIER + <i>and</i> + CLASSIFIER | 133 |
| Table 62: 'Listed people' | 134 |
| Table 63: LOCATION + <i>and</i> + LOCATION | 134 |
| Table 64: ARGUMENT + <i>and</i> + ARGUMENT | 135 |
| Table 65: PHENOMENON/PHENOMENA + <i>and</i> + ITS/THEIR CHARACTERISTICS | 135 |
| Table 66: PHENOMENON/PHENOMENA + <i>and</i> + PHENOMENON/PHENOMENA ... | 136 |
| Table 67: Difference between 'Listed processes' and 'Addition' motifs for <i>and</i> | 137 |
| Table 68: CAUSE + <i>and</i> + SIGNAL OF CONSEQUENCE + CONSEQUENCE | 138 |
| Table 69: Phraseologies with <i>between</i> | 138 |
| Table 70: Semantic sequences and motifs for <i>and</i> – frequency and percentage in pooled corpus and disciplinary/institutional sub-corpora | 139 |
| Table 71: Semantic sequences and motifs for <i>and</i> – frequency and percentage in pooled corpus and disciplinary/institutional sub-corpora – adjusted percentages | 140 |
| Table 72: Semantic categories and motifs for <i>that</i> | 142 |
| Table 73: 'The voice of others' (1)..... | 143 |
| Table 74: 'The voice of others' (2)..... | 144 |
| Table 75: 'The voice of others' (3) | 144 |
| Table 76: 'The voice of others' (4) | 145 |
| Table 77: 'The voice of others' (5) | 145 |
| Table 78: 'The voice of the essay writer - arguing' (1)..... | 147 |
| Table 79: ESSAY-WRITER ANGLE ON STATEMENT element compared with CHARACTER OF STATEMENT element | 148 |
| Table 80: 'The voice of the essay writer - arguing' (2)..... | 148 |
| Table 81: 'The voice of the essay writer - arguing' (3)..... | 149 |
| Table 82: 'The voice of the essay writer – interpretation' | 149 |
| Table 83: PHENOMENON + <i>that</i> + DETAIL | 150 |
| Table 84: Ellipsis of <i>the fact that</i> to 'That' | 150 |
| Table 85: <i>the fact that</i> used for Cause-Effect and Evidence-Interpretation functions | 151 |
| Table 86: <i>the fact that</i> used in expression of advantage, reason and problem | 152 |
| Table 87: 'The voice of others' – frequency across disciplinary/institutional sub-corpora | 152 |
| Table 88: 'The voice of the essay writer' – frequency across disciplinary/institutional sub-corpora | 152 |

| | |
|---|-----|
| Table 89: Non 'voice' phraseologies for that - frequency across sub-corpora | 153 |
| Table 90: Comparison of first and second sample frequencies for I/we + ESSAY- WRITER ANGLE ON STATEMENT + <i>that</i> + STATEMENT | 154 |
| Table 91: Comparison of first and second sample frequencies for PHENOMENON AS EVIDENCE + INTERPRETIVE ANGLE + <i>that</i> + INTERPRETATION | 154 |
| Table 92: Comparison of first and second sample frequencies for the fact that phraseologies | 154 |
| Table 93: Semantic analysis of <i>as</i> in pooled sample | 155 |
| Table 94: Sequencing possibilities for the semantic motif 'interpretation' | 156 |
| Table 95: An illustration of sequencing possibilities for the semantic motif 'Interpretation: perception' | 156 |
| Table 96: Examples of 'Interpretation: perception' | 157 |
| Table 97: Examples of 'Interpretation: portrayal' (1) | 157 |
| Table 98: Examples of 'Interpretation: portrayal' (2) | 158 |
| Table 99: Examples of 'Interpretation: portrayal' (3) | 158 |
| Table 100: Examples of 'Interpretation: judgement' | 159 |
| Table 101: Examples of 'Interpretation: classification' (1) | 159 |
| Table 102: Example of 'Interpretation: classification' (2) | 159 |
| Table 103: 'Reason-Result' semantic sequences with <i>as</i> | 160 |
| Table 104: 'Reason-result' semantic sequences with <i>as a result</i> or <i>as a result of</i> .. | 161 |
| Table 105: 'Exemplification' with <i>such as</i> | 161 |
| Table 106: Semantic sequences with <i>as well as</i> for 'addition' | 162 |
| Table 107: Semantic sequences for 'Attribution' | 163 |
| Table 108: Head nouns' following <i>as</i> in 'Function' phraseologies | 164 |
| Table 109: <i>bastion</i> used as <i>source</i> | 164 |
| Table 110: Sequencing possibilities for the semantic motif 'Function' | 165 |
| Table 111: Semantic sequences for 'Function' | 165 |
| Table 112: Sequencing possibilities for the semantic motif 'Role' | 165 |
| Table 113: Examples of 'Role' (1) | 166 |
| Table 114: Examples of 'Role' (2) | 166 |
| Table 115: Semantic sequences for 'Role' | 166 |
| Table 116: Sequencing possibilities for the semantic motif 'Framing or Conceptualisation' | 167 |
| Table 117: 'Similarity' sequences | 167 |
| Table 118: 'Ambiguous' lines which are categorised as 'similarity' rather than 'attribution' | 168 |
| Table 119: Semantic sequences for 'Emphasis of extent' | 168 |
| Table 120: Emphasis/intensification realised with <i>as much as</i> | 169 |
| Table 121: Semantic sequences for 'Simultaneity' | 170 |
| Table 122: Frequency for semantic groupings with <i>as</i> across disciplinary/institutional sub-corpora | 170 |

| | |
|--|-----|
| Table 123: Semantic analysis of <i>this</i> in pooled sample..... | 175 |
| Table 124: 'Encapsulation as scholarly angle' (1) | 177 |
| Table 125: 'Encapsulation as scholarly angle' (2) | 177 |
| Table 126: 'Encapsulation as scholarly angle' (3) | 177 |
| Table 127: 'Encapsulation as scholarly angle' (4) | 178 |
| Table 128: 'Encapsulation as object' (1) | 178 |
| Table 129: 'Encapsulation as object' (2) | 179 |
| Table 130: 'Encapsulation as process' | 179 |
| Table 131: Examination of co-text to confirm membership within grouping of problematic 'process' lines | 180 |
| Table 132: <i>this</i> + PART/WHOLE OF ESSAY TEXT ITSELF..... | 181 |
| Table 133: <i>this</i> + SITUATION OR CIRCUMSTANCE (1) | 181 |
| Table 134: <i>this</i> + SITUATION OR CIRCUMSTANCE (2) | 182 |
| Table 135: <i>this</i> + SITUATION OR CIRCUMSTANCE (3) | 182 |
| Table 136: <i>this</i> + SENSE | 182 |
| Table 137: <i>this</i> + EVALUATION (1)..... | 183 |
| Table 138: <i>this</i> + EVALUATION (2)..... | 184 |
| Table 139: <i>this</i> + INTERPRETATION | 184 |
| Table 140: <i>this</i> + INTERPRETATION (RELATIONSHIP TO OTHER IDEAS)..... | 185 |
| Table 141: <i>this</i> + ADDITION | 185 |
| Table 142: <i>this</i> + EFFECT (1) | 185 |
| Table 143: <i>this</i> + EFFECT (2) | 186 |
| Table 144: <i>this</i> + EXEMPLIFICATION..... | 186 |
| Table 145: EXEMPLIFICATION PHRASE [INCLUDING <i>this</i>] + EXAMPLE..... | 186 |
| Table 146: <i>this</i> + CAUSE | 187 |
| Table 147: <i>at/to/till</i> + <i>this</i> + POINT-IN-TIME ENCAPSULATION | 187 |
| Table 148: <i>in</i> + <i>this</i> + MANNER | 188 |
| Table 149: Frequencies of <i>this</i> phraseologies across disciplinary/institutional sub- corpora | 189 |
| Table 150: Comparison of first and second sample frequencies for 'Encapsulation as scholarly angle' | 190 |
| Table 151: Comparison of first and second sample frequencies for 'Encapsulation as object' | 190 |
| Table 152: Comparison of first and second sample frequencies for 'Encapsulation as the essay text' | 190 |
| Table 153: Comparison of first and second sample frequencies for 'Meaning of preceding statement' | 190 |
| Table 154: Comparison of first and second sample frequencies for 'Addition to preceding statement' | 191 |
| Table 155: Comparison of first and second sample frequencies for 'Exemplification of preceding statement' | 191 |

| | |
|---|-----|
| Table 156: Comparison of first and second sample frequencies for ‘Cause of preceding statement’ | 191 |
| Table 157: Patterns of continuity in the semantic motifs or semantics sequences of <i>of, and, that, as</i> and <i>this</i> | 196 |
| Table 158: Common features of student ‘Humanities and Social Sciences’ texts (Durrant (2015)) | 198 |
| Table 159: PROPERTY + <i>of</i> + PHENOMENON sequences from the pooled sample . | 201 |
| Table 160: Disciplinary differences in the phraseologies of <i>this</i> | 215 |

Acknowledgements

Firstly, I would like to acknowledge my supervisors, Dr Sue Wharton and Dr Malcolm MacDonald, for their support, encouragement and for the consistently wise and practical advice which has guided this project.

I would also like to acknowledge my previous employers at Royal Holloway University for support in terms of financial contributions to tuition fees as well as provision of study leave, and my current employers at the University of Reading for generous provision of study leave around Christmas 2015 without which I would have struggled to complete this thesis within the stipulated time frame.

I would like to thank Dr Duncan Hunter for encouraging me to undertake a PhD thesis.

Finally, I would like to thank my partner Mike for his patience, his emotional and technical support, and for enduring a very large number of pretty dull weekends over the past five and a half years.

Declaration

I declare that the present thesis has been researched and composed by myself and has not been submitted elsewhere for any other degree or professional qualification.

Karin Whiteside

Abstract

This thesis involves a comparative lexico-grammatical analysis of third-year student writing belonging to the Essay genre family (Nesi and Gardner, 2012) in two disciplines, History and PIR (Politics and International Relations), from two UK higher educational institutions. The project adopts a corpus-driven approach which was developed by Groom (2007) in his analysis of professional academic writing in Literature and History: statistically significant grammatical words are identified using a keyness analysis, and the phraseological patterning around these grammatical keywords is then qualitatively analysed and phraseologies are categorised according to their semantic purposes. In the project five grammatical keywords - *of*, *and*, *that*, *as* and *this* - were analysed across four sub-corpora each consisting of student writing from one of the two disciplines at one of the two institutions. It was found that there were more similarities than there were variations in the semantic patterning of grammatical keywords across the four disciplinary/institutional sub-corpora, and that these similarities could to a large extent be explained in terms of the shared features of student Humanities and Social Sciences writing (Durrant, 2015). The variations that occurred fell along disciplinary rather than institutional lines and it is argued that, with regards to both similarities and differences, in the case of these two disciplines at the two target institutions, discipline seems to override institution as an influence at lexico-grammatical level on the nature of student academic writing. It is also argued that Groom's (2007) approach is an extremely useful one to take in analysis of student writing as it uncovers lexico-grammatical features which occur extremely regularly within student texts and thus, from a pedagogical perspective, are of high value in terms of how much of the text they 'operationalize' (Bruce, 2011, p. 6).

Chapter 1: Introduction

This project takes a corpus-driven approach in the bottom-up analysis of the lexicogrammatical features of good third-year writing in the disciplines of History and Politics/International Relations (PIR) at two UK Higher Education institutions. The motivation for the project arose from questions I began to ask myself about the efficacy of the teaching of academic writing in my role as an EAP tutor at a UK higher education institution. The questions I asked followed two related themes.

Firstly, I felt increasingly that the type of knowledge imparted about 'how to write an essay' in 'generic' (as opposed to discipline-specific) EAP Pre-sessional and In-session programmes was limited in its effectiveness, and that there was potentially a lot more that could be achieved in academic writing instruction. Such courses equip students with the ability to emulate reasonably effectively the surface conventions of essay writing, the ability to structure introductions and conclusions, structure paragraphs, write topic sentences, and to adhere to rules about register such as use of formal vocabulary and avoidance of too many personal pronouns. However, there is another level to academic writing which is arguably much more important for students in order for them to write successfully within their degree courses. This level is linked to making a convincing argument, demonstrating depth in terms of content understanding, and showing criticality, all factors that potentially have a much stronger influence than surface features on how positively a student will be assessed within her/his studies. To some extent this level can be beyond the remit of an EAP or writing instructor in that it is bound up with content understanding imparted by subject specialists. However, it is my belief that, in terms of how this level manifests itself in particular features of writing, in expectations about language and discourse, there is a role for the writing/EAP instructor in making its elements more transparent for students.

Secondly, and related to this first concern, is the question of the influence of discipline on what constitutes a good piece of student writing. Tasked in 2009 with the job of developing a discipline-specific EAP writing course for international

students embarking on taught master's degree programmes in Politics and International Relations, I sought out the literature which could inform my approach on possible particularities of an 'essay' in this disciplinary field. At that point in time, three years prior to the publication of Nesi and Gardner's (2012) survey of student genres across disciplines at UK universities, very little research existed which actually examined student-produced texts – most studies looked at tasks – and very little research compared student texts from a disciplinary perspective. I believed I needed to know more about the extent to which in disciplines for which, for example, the primary genre is the essay, students were writing the same type of text, or, conversely, the extent to which surface similarities in macro-structure - introductions, conclusions, paragraphing - might be masking significant disciplinary differences in language and discourse at another level in the text. Another possibility which sits between these two extremes was that there might be clusters of disciplines, 'meta-disciplinary' groupings in which essays share lexicogrammatical characteristics. A further possibility is that differences in student writing were not in fact wholly disciplinary and might be explained by institutional influences, for example, the culture of a particular institution or department, or the preferences and particular requirements of individual tutors.

There is considerable depth and breadth to the knowledge we have about disciplinary characteristics of *professional* academic writing as the result of a wide range of research conducted over more than three decades; Bazerman's (1981) and Becher's (1987) work can be seen as the beginning of this research tradition. With the advent of corpus approaches, research into disciplinary features of professional academic writing has arguably been able to draw more reliable and comprehensive conclusions dealing as it does with much larger samples of target texts. Hyland's highly influential work (e.g., 2000, 2001, 2002a, 2005, 2007, 2008, 2009) has contrasted writing across the four categories of Becher and Trowler's (2001) disciplinary matrix. Other studies have involved closer analysis and comparison of writing in a smaller number of particular disciplines, often from very different disciplinary areas (e.g. Conrad, 2001; Cortes, 2004; Charles, 2006; Silver, 2006), but also from disciplines that sit relatively closely within Becher and Trowler's (2001)

disciplinary matrix (e.g. Bondi, 2004, 2006; Bondi and Silver, 2004; Groom, 2005, 2007; Nishina, 2010; Malavassi and Mazzi, 2010). Nishina makes a persuasive argument for the usefulness of the latter focus arguing (2010, p. 13):

it could be objected that we can easily predict the existence of differences between language used in soft and hard disciplines even before undertaking in-depth research. Differences within the same epistemological field (i.e hard or soft) are harder to predict In-depth studies of epistemologically similar disciplines is valuable not only in that it contributes to our understanding of disciplinary differences, but also because it presents a more stringent test for disciplinary differences.

Of particular interest in terms of analysis of disciplinary features of two 'close' disciplines is Groom's (2007) study of the phraseology of professional academic writing in History and Literary Criticism. This research takes a bottom-up approach to the qualitative analysis of semantic patterning (Hunston, 2008) around grammatical keywords in corpora of research articles in the two disciplines. The justification for such an approach is the argument that the phraseological patterning around grammatical words is highly revealing of semantic preferences of particular disciplinary discourses (Hunston, 2008; Groom, 2010).

In comparison to professional academic writing, the body of research concerning disciplinary characteristics of student academic writing is, so far, much less substantial, although student writing has received considerably more attention in the past few years, particularly with the completion of the BAWE corpus project (Nesi and Gardner, 2012). Nesi and Gardner's (2012) work provides a comprehensive analysis of the typology of student academic written genres and their distribution across sixteen disciplines. However, Nesi and Gardner's analysis of linguistic features of student genres is less comprehensive in its coverage; a corpus analysis using Biber's (1988) Multi-Dimensional (MD) framework is carried out at the level of the four disciplinary groupings of BAWE rather than at the level of the sixteen specific disciplines. At disciplinary level the focus of Nesi and Gardner's linguistic analyses are selective rather than systematic.

At the beginning of this project, between 2011 and 2012, a review of other studies which investigated disciplinary features of student texts showed there were significant research gaps. Although some studies had contrasted the linguistic features of particular disciplines at lexico-grammatical level (e.g. Cortes, 2004; Thompson, 2009; Li and Wharton, 2012), these studies had largely applied pre-existing linguistic frameworks often formulated in the study of professional academic writing. Since that time, however, two large bottom-up cross-disciplinary studies of the lexico-grammatical features of student writing have been undertaken and published. Hardy and Römer (2013) have conducted an Multi-Dimensional (MD) analysis of student writing in the sixteen disciplines captured in the Michigan Corpus of Upper Level Student papers (MUCUSP, 2009) identifying four functional dimensions for this corpus finding both similarities and differences linked to discipline and genre. Durrant (2015) has investigated disciplinary variation within the BAWE corpus in a study which did not assume pre-determined disciplinary categories but let groupings emerge through analysis of variations in patterns of lexical bundle use. Four main disciplinary groupings were revealed by this analysis, science/technology, humanities/social sciences, life sciences and commerce. These studies represent considerable progress in the mapping of disciplinary similarities and differences in academic writing across the academy from a purely student angle. There is, nonetheless, still considerable room for closer, more in-depth studies of pairs or small numbers of disciplines and for more bottom-up studies of disciplinary features of student texts that allow for completely new features of these texts to be identified.

Research which frames student writing in purely disciplinary terms, however, needs, to an extent, to be problematised on the grounds of the peculiarities of student writing. Such factors as the low status and consequent careless labelling of genres (Johns, 2008) and the communicative purpose of accreditation (Nesi and Gardner, 2012) clearly separate student writing from parallel professional disciplinary activity. The argument that the student journey through their undergraduate degree serves as an 'apprenticeship' into the disciplinary community is also contested (e.g. Berekenkotter and Huckin, 1995; Candlin and Plum, 1999;

Woodward-Kron, 2004). Influences other than discipline, such as those from potentially conflicting social or cultural identities (e.g. Russell, 1997; Lillis, 1997; Ivanic, 1998), and from particularities of at the level of institution or individual tutor (e.g. Lea and Street, 1998), need to be taken into account. Studies of disciplinary characteristics of student writing have generally not, however, factored non-disciplinary influences into their research design; Li and Wharton's (2012) research is one exception to this generalisation as in their study student writing was contrasted not only across disciplinary fields but also across institutions.

My project seeks to uncover potential disciplinary characteristics of student writing at lexico-grammatical level in two close disciplines/degree programmes, History, and Politics/International Relations. These two areas sit closely on the disciplinary spectrum; in his analysis of lexical bundles in BAWE texts Durrant (2015) found that they form a close cluster, along with Law, within a broader humanities and social sciences grouping. They thus afford, as Nishina (2010, p. 13) argues, an arguably 'more stringent test for disciplinary differences'. It is hoped that the analysis may identify lexico-grammatical features which are characteristic of successful student writing beyond those macro-structural and stylistic features which have more traditionally been the central focus of EAP writing instruction. With this goal in mind, an approach similar to the one taken by Groom (2007) in his contrastive analysis of professional disciplinary writing is particularly attractive if one accepts Hunston's (2008) and Groom's (2010) contention that analysing the semantic patterning around key grammatical words is an effective way to get at phraseological and semantic preferences of particular disciplinary discourses. Such an approach is also attractive as it provides a bottom up lexico-grammatical analysis of student writing. It thus contributes to an area of student academic writing research in which there is still considerable room for further work. Taking into account also the need argued in the paragraph above for accounting for other potentially strong influences on the nature of student writing, the research design incorporates both a cross-disciplinary and cross-institutional analysis of four sub-corpora of texts representing writing completed in the two subject areas in two different UK HE institutions.

To summarise, this project aims to achieve the following objectives:

- To reveal the extent to which there are similarities or differences to be found at the lexico-grammatical level in good student History and Politics/International Relations writing at two UK HE institutions.
- To assess how far any similarities or differences which are found can be accounted for by the influence of 'discipline' or by the influence of 'institution'.
- In doing the above, to uncover within the target texts lexico-grammatical features which may have pedagogical value for an EAP teaching context.
- Also, in doing the above, to consider the usefulness of Groom's (2007) approach when it is applied to the analysis of student, as opposed to professional, written academic discourses.

Chapter 2: Literature Review

This project involves a contrastive discourse analysis of the language of specialised texts and is for this reason strongly underpinned by the idea in applied linguistics of 'genre', an abstract theoretical construct used for grouping and explaining texts by what they share in terms of cultural context, social purpose, structure and language. A basic assumption underlying my research is the need to account for the influence of genre when examining discourse produced in academic institutions. I therefore begin my review of the literature with a discussion of the main perspectives on genre, which, although overlapping significantly, still differ crucially with regards to the centrality of language features to genre research. The approach followed in my project assumes the usefulness of a close and central focus on language for adding to genre knowledge about student academic writing.

Following the discussion of genre, I examine the concepts of 'discipline' and 'disciplinary discourse', firstly with regards to their primary relationship to professional academic writing, and then as they relate to the more peripheral phenomenon of student writing occurring within degree programmes grouped within disciplines in the higher education system. I argue that with respect to both the particular nature of and the competing influences on student writing, concepts of discipline and disciplinary discourse become potentially more problematic. For this reason it cannot be automatically assumed that discipline explains discourse features of student writing, and, therefore, other potential influences need to be taken into account in investigations of student writing within disciplines.

I start the final section of my review of the literature with a discussion of key issues and debates related to the role of corpus tools and approaches in discourse analysis. I then focus particularly on an analytical tool employed in my study, that of a keyword analysis as the starting point for further close analysis of concordance samples from specialised corpora, and the implications of choices made from a keywords list when investigating potentially 'disciplinary' features in specialised

texts. I examine the case made for the choice of 'grammatical' words in this context, which is the approach I have adopted for my study.

2.1. Genre

Research into academic writing is informed by a concept of 'genre' which took shape from the late 1980s when a shift occurred in the approach to writing in terms of both theory and pedagogy from 'psycholinguistic/cognitive literacy theories and "The Process Approach"' to 'a contextual approach, to analyses of the situations in which writing takes place' (Johns, 2002, p. 3). There is not, however, a unified overarching conceptual framework within which contemporary genre research takes place; rather, it is widely agreed that there are three major 'traditions' (Hyon, 1996; Johns, 2008), work informed by Systemic Functional Linguistics (SFL) in 'the Sydney School', work within North American New Rhetoric (NR) studies, and work within English for Specific Purposes (ESP). These traditions are seen as 'distinct – yet often overlapping – theoretical frameworks' (Tardy, 2011, p 2). They share an emphasis on the social nature of writing, the idea that writing is shaped much more by the social and cultural context within which it takes place than by the internal workings of the individual writer's mind (an idea associated with earlier Process writing pedagogies). The three traditions also broadly share the view that 'genre' is an abstract resource for both constructing and explaining texts as responses to repeated social situations within particular cultural contexts. Their differences lie largely in the different scholarly goals they prioritise, respectively, to enrich the theorisation of social activity in the case of NR, to enrich the theorisation of language in order to develop improved writing pedagogies in the case of the Sydney School, or to enhance understanding of particular communicative purposes in specific work or study contexts in order to develop improved writing pedagogies which is the case for ESP.

2.1.1. New Rhetoric and genre

In NR research genre is seen as social action and the theoretical focus is on understanding this action; Barwarshi and Reiff (2010, p. 59) argue that in what they term Rhetorical Genre Studies (RGS) 'understanding contexts (and their performance) is both the starting point for genre analysis and its goal'. Two scholars, Carolyn Miller and Charles Bazerman, have been highly influential in shaping the concept of genre within New Rhetoric scholarship. In her seminal 1984 article Miller argues 'a rhetorically sound definition of genre must be centred not on the substance or form of the discourse but on the action it is used to accomplish' (p. 151) and defines genres as 'typified rhetorical actions based in recurrent situations' (p. 159). Bazerman argues that rather than seeing genres as textual forms, they should be seen as 'forms of life, ways of being, frames for social action' (Bazerman, 1994, p. 1) (Swales points out Bazerman's 'frame' is 'subtly different from Carolyn Miller's ... definition centred on the accomplished rhetorical action itself' (2009, p. 6)).

A number of NR scholars focus on how genre behaves with respect to time. Berkenkotter and Huckin (1993, p. 481) argue that to effectively function over time, genres 'must accommodate both stability and change' and Shryer (1994, p. 108) describes genres as 'stabilised-for-now or stabilised-enough sites of social and ideological action'. Devitt (1993, p. 577) adds that '[g]enre not only responds to, but also constructs recurring situations'. Russell sees genre as the 'operationalization' of Cole and Engestrom's 'activity system', object/motive directed human interaction mediated by 'tools' such as 'machines, writing, speaking, gesture, architecture, music etc.' (1993, in Russell, 1977, p. 510) and maintains that (p. 513):

As "forms of life", genres and the activity systems they operationalize (temporarily) are regularized or stabilized through routinized tool use within and among (sub)groups. This context is an ongoing accomplishment, not a container for actions or texts. The behaviour of individual writers/subjects is

constantly recreated through the specific actions of people together. Thus, genres are always only “stabilized for now” in Shryer’s (1993) apt phrase.

Genre analysis informed by NR theory generally takes an ethnographic rather than linguistic orientation ‘offering thick descriptions of academic and professional contexts surrounding genres and the actions texts perform in these situations’ (Hyon, 1996, p. 696). An example of recent research in this tradition is Miller and Shepherd’s ‘ethno-methodological’ examination of the ‘emergent’ genre of weblogs in academic settings (2004). Another example is a study by Bawarshi and Reiff (2011) investigating pre-existing genre knowledge of students on entry to first-year college composition courses in the US and how this knowledge either aids or hinders students’ accessing of academic discourse; this involved multiple research methods including surveys about past literacy experiences, interviews with students about the texts they produce, and analysis of both writing produced on first-year composition courses, and the syllabi and assignment documents.

2.1.2. SFL and genre

The fundamental difference between NR and SFL approaches to genre is the privileging of context in the case of the former, and the privileging of text in the case of the latter. This can be seen very clearly when Miller’s and Martin’s words are contrasted:

[w]hat we learn when we learn a genre is not just a pattern of form or even a method of achieving our own ends. We learn, more importantly, what ends we may have (Miller, 1984, p. 38)

[d]iscourses are tools – they do things. That is why they have evolved and thus their functionality determines their character (Martin, 1993, p. 221).

For NR scholars, studying genre uncovers social activities and intentions, whilst the linguistic ‘pattern of form’ of these activities is of lesser importance. For SFL

scholars, language very much *embodies* the activity and therefore the 'character' of language is the central focus of research.

In the SFL approach genre can be seen as an extension of the systemic functional linguistic framework of language and social context, a 'rich conception of language as a meaning making system' (Martin, 2009, p. 11) developed by Halliday (1978) in which grammar is a resource for meaning-making and texts embody socially-situated semantic choices. Halliday used *register*, 'the clustering of semantic features according to situation types' (p. 68), rather than genre as his central analytical construct, register being determined by the three elements of social setting, 'field', the activity taking place, 'tenor', how participants socially relate, 'mode', the particular channel by which communication is taking place (p. 33). Martin (1984) developed a theory of genre from the systemic functional framework defining genre as 'a staged, goal-oriented, purposeful activity in which speakers engage as members of a culture' (Martin, 1984, p. 25). He states the following about the 'job' of genre (2009, pp. 12-13):

... it was to coordinate resources, to specify just how a given culture organizes this meaning potential into recurrent configurations of meaning, and phases of meaning through stages in each genre. The basic idea here is that we cannot achieve our social purposes all at once, but have to move in steps, assembling meaning as we go ... the high-level position of genre in the model provided a way of talking holistically about the social purposes of texts and the ways in which different genres marshalled different resources to achieve their goals ... one of the reasons for separating genre from field, tenor and mode was to allow for shifts in [these] variables from one stage of a genre to another.

Whereas register operates at the contextual level of 'situation', Martin sees genre as operating at the overarching contextual level of 'culture' (1997, p. 37; 2009, p. 12) stating that (2009, p. 13):

... out of all the things we might do with language, each culture chooses just a few, and enacts them over and over again ... slowly adding to the repertoire as needs arise, and slowly dropping things that are not much use. Genre theory is thus a theory of the borders of our social world, and our familiarity with what to expect.

Genre analysis informed by SFL theory focuses more on the text than the context often starting with an investigation of social purpose as evident from the generic stages in a text moving from this to register analysis, then to how register is realised linguistically. Recent examples of SFL-informed research into academic writing include work by Woodward-Kron (2009), Donahue (2012) and Nesi and Gardner (2012).

Woodward-Kron (2009) undertakes a longitudinal study of the lexico-grammatical development of Education students' writing during the course of their undergraduate degree programme. Her study looks at the development of lexico-grammatical resource for reasoning and explaining, reporting knowledge claims and implicit and explicit forms of engagement, and she finds differences in the choices of and frequency of use of these resources between first-year and third-year texts and between lower-scoring and higher-scoring writers. For example, for reporting knowledge claims, one finding was an increased use of 'Abstract Participant with human Deictic' (e.g. *Sternberg's triarchic theory*) between first and third year texts as well as greater use of this form of reporting in higher-scoring texts (p174).

Donahue (2012) analyses the 'taxonomic film analysis', a key written genre for film studies students. He identifies the linguistic elements expressing SFL experiential and textual 'metafunctions' (related respectively to 'field' and 'mode' components of SFL register) and contrasts their use in texts by two different first-year student authors, one who successfully realises the taxonomic film analysis genre, and one who does not. He argues that the two students 'are using language differently because they are doing different things: they are analysing differently' (p. 10), and that this demonstrates the closeness of the relationship between language use and purposes in student writing in the context of a particular discipline (p. 15).

Nesi and Gardner (2012) make a very significant contribution to understanding of the range of genres written by undergraduate and postgraduate taught master's students across disciplines at UK universities the findings of which will be discussed in greater detail later in this chapter. They draw on SFL genre analysis to 'identify

the educational purposes and stages that typify and distinguish genre families' (2012, p. 11) in describing student genres across disciplines within the UK higher education context. In the process of categorising student texts they posed questions regarding 'function, stages and purpose' (p. 34) to identify thirteen genre families that can be grouped under five broad functions (pp. 32-43). These genre families differ from Martin's SFL classifications, developed to explain primary and secondary school genres, partly because they 'aimed to develop them by grouping and regrouping similar assignments, rather than imposing a classification developed for other contexts', and partly due to the influence of frameworks outside of SFL, Swales' work in the ESP tradition of genre analysis (see below) and work in the academic literacies field (2012, p. 11).

2.1.3. ESP and genre

In contrast to SFL and NR approaches which are both strongly theoretically-based traditions, the ESP approach to genre has from its inception been 'applied' in that its starting point was a practical concern, that of specific-purpose language instruction. This approach to genre 'bridges linguistic and rhetorical traditions' (Bawarshi and Reiff, 2010, p. 41) with ESP's principal theorist, Swales, drawing from both SFL and NR traditions in *Genre analysis: English in Academic and Research Settings* which established the theoretical and methodological basis for ESP genre analysis (1990, pp. 40-44)). Swales' definition of genre is as follows (1990, p. 58):

A genre comprises of a class of communicative events, the members of which share the same communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style. Communicative purpose is both a privileged criterion and one that operates to keep the scope of the genre as here conceived narrowly focused on comparable rhetorical action. In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience.

Two key concepts in Swales' definition of genre are 'discourse community' and 'communicative purpose'. Discourse communities, as conceived by Swales, are 'socio-rhetorical networks that form in order to work towards sets of common goals' and a characteristic of established members of a discourse community is 'familiarity with particular genres used in communicative furtherance of set goals' (p. 9). Thus, genres are the property of the discourse community rather than the individual, any other grouping, or a wider speech community (p. 9). Swales argues that the common thread of communicative purpose binds discourse community activity, genre identity, and task (p. 10) and in his 1990 work the '*privileged* property of a genre' was conferred on communicative purpose overriding 'form, structure, and audience expectations' in determining how 'prototypical' it was (p. 52).

Swales has since somewhat revised his view about the role of communicative purpose. Askehave and Swales (2001) argue that with the increased number of genre studies and consequent greater understanding of the complexity of the concept of genre 'the concept of 'communicative purpose' has also become more complex, multiple, variable, and generally hard to get at' (p. 195) and suggest abandoning establishing communicative purpose as the first step in genre analysis for new step-by-step procedures, either linguistic or ethnographic, in which first impressions about purpose are revisited at a later 'repurposing' step on the analysis process (pp. 207-208).

Swales has also revised his conception of genre. In his 2004 work Swales abandoned definition as a means to explain genre in favour of 'metaphor' arguing that definitions tend not to be true 'in all possible worlds and all possible times' and that they can 'prevent us from seeing newly explored or newly emerging genres for what they are' (2004, p. 61). His new model of genre consists of six metaphors, including 'Frames of Social Action' (drawn from Bazerman, 1997, p. 19) as metaphor for 'Guiding Principles', 'Biological species' as metaphor for 'Complex Historicities', and

‘Institution’ for ‘Shaping Contexts; Roles’ (2009, p. 6). However, Swales has suggested that this revision was to an extent ‘forced’ (2009, p. 5):

I concluded that I could not basically reiterate a position espoused more than a decade previously but, true to the grand academic imperative, would have to offer something new. (Even though, in my heart of hearts, I felt there was little actually wrong with that old earlier characterization, except for a mistaken emphasis on genres as distinct independent entities).

ESP genre analysis tends to move from the context, the genre and communicative purpose as understood and identified by the discourse community, to the text. Textual analysis then generally starts at the level of schematic structure and works from this down to the level of lexico-grammatical features. This is exemplified in Swales’ analysis of research articles (RAs) (1990, pp. 140-150). He identifies the three ‘move’ structure of introductions, the CARS (Creating a Research Space) structure - a ‘move’ is defined by Swales and Feak (1994, p. 35) as a ‘bounded communicative act that is designed to achieve one main communicative object’. Swales then identifies the ‘steps’ that can occur within each move, such as ‘Counter-claiming’ or ‘Question-raising’ in the second move of the structure. Finally, analysis is undertaken of the linguistic exponents that signal steps, the nature of citation, and the use of reporting verbs. At this lexico-grammatical level, in comparison to SFL, ESP scholarship tends to take a top-down, selective approach to what linguistic features come in for close investigation. Recent examples of research in the ESP tradition include Samraj’s (2008) examination of master’s theses, both in contrast to RAs and in terms of disciplinary differences, in which overall organization, moves within introductions, and citation patterns were studied, and Basturkmen’s (2012) examination of discussion sections in Dentistry RAs, in comparison to those from Applied Linguistics, involving ‘analysis of steps, sub-steps and sequences within moves’ (p. 134).

2.1.4. Paltridge and Bruce

Two scholars particularly interested in genre in the context of academic writing, Paltridge (1994; 1995a; 1995b; 1996; 2002) and Bruce (2008; 2009), make contributions to genre scholarship from an EAP perspective offering solutions to what they view as gaps in both SFL and ESP explanations of genre. In their work they both draw on Biber's (1989) argument regarding the distinction between genre and text type. Biber develops a typology of texts according to five dimensions composed of features of lexis and syntax that co-occur frequently, and argues that this typology demonstrates that there is not a straightforward relationship between genre and text type; he argues that genres are only valid when defined and distinguished according to 'systematic nonlinguistic criteria', whereas text types are distinguished via 'strictly linguistic criteria (similarities in the use of co-occurring linguistic features)'. Both Paltridge's and Bruce's research supports this view, and both offer different but related genre models which seek to bridge the gap between genre and language features. Both scholars also draw on the work of both Hoey (1983) and Crombie (1985) to redress problems they have identified in the SFL and ESP frameworks.

Paltridge examined the criteria used to identify textual boundaries in examples of genre analysis and finds that these were non-linguistic (1994, p. 288). He argues that divisions in texts are perceived more cognitively than linguistically (p. 295) and that many genre analysts have tried to stretch their models beyond the limits of their applicability by trying to explain certain 'textual aspects of language' grammatically when these are better accounted for from a pragmatic perspective (ibid.). Paltridge argues that the genre analysis models of SFL scholars and Swales do not fully account for 'cognitive' as well as 'social' linguistic features (1995a, p. 393) and he offers a framework to redress this gap. He argues that genres are identified by the social elements of his framework, on the basis of pragmatic and perceptual rather than linguistic characteristics of communicative events (ibid.) and identifies layers of 'discourse structures' within examples of genres, from the macrostructure of texts down through, drawing on the work of Hoey (1983) and

Crombie (1985), discourse elements (e.g. 'situation'), discourse relations (e.g. 'situation-problem') to semantic relations, which show the relationship between propositions in a text (e.g. 'statement-exemplification') and occur within or between discourse elements (1995a, 1995b). He argues that at the level of discourse what typifies a genre is not necessarily any one particular single element of discourse structure being present, but, instead, occurrence together and interaction of a number of discourse structure levels and elements which can be realized by a vast range of linguistic choices meaning it is impossible to accurately predetermine specific language choices in a particular instance of a genre (1995a, p. 403). In terms of pedagogical implications, he argues that genre analyses in which the notions of genre and text type are conflated or in which only one or the other is presented risk not providing students with 'a complete view of the discourse components of texts' (1996, p. 240) and that therefore how text types and genres relate should be addressed in language and EAP classrooms (1996, 2002).

Bruce, like Paltridge, distinguishes between social and cognitive elements in his approach to genre (2008, 2009). He uses Biber's (1989) conception of text types in the cognitive component of his framework. He defines 'social genre' as 'socially recognized constructs according to which whole texts are classified in terms of their overall purpose' (2009, p. 106) and involve knowledge of context, epistemology, writer stance (such as Hyland's (2005) conception of metadiscourse), and content schemata (p. 107). He defines 'cognitive genre' as 'the overall orientation and internal organisation of a segment of writing that realizes a single, more general rhetorical purpose to represent one type of information within a discourse' (p. 107). These segments are usually combined in the creation of whole texts. Bruce proposes a model consisting of four 'cognitive genres' which he argues occur frequently in academic English writing drawn from the four text types found by Biber (1989, p. 39) to be most frequent in academic prose (ibid.). The four cognitive genres are 'report', 'explanation', 'discussion' and 'recount' and in modelling these Bruce identifies 'rhetorical focus', 'gestalt pattern', 'discourse pattern' (drawing on the work of Hoey) and 'interpropositional relations' (drawing on the work of Crombie)(p. 108). Bruce contends that his cognitive genre framework redresses

issues in the practical application of genre approaches in the EAP context. Regarding New rhetoric and SFL approaches respectively, he argues (*ibid.*, p. 12):

while both the contextual and textual approaches account for important dimensions of genre knowledge, neither provides a comprehensive approach to operationalizing discourse competence in relation to the writing of academic genres, and, crucially, both ignore the important mediating domain of metacognitive, procedural knowledge, [as] operationalized ... in terms of the elements of the cognitive genre model.

Both Paltridge and Bruce could be criticized for offering somewhat ‘over-engineered’ solutions to the issues they have identified with the efficacy of SFL and ESP models in an EAP context. Their solutions entail potentially unwieldy models involving the application of further layers of textual analysis drawn from a number of pre-existing frameworks. From the perspective of the EAP teaching of student writing, a further possible criticism is that these frameworks have not emerged from actual analysis of student texts meaning there is a ‘top-down’ rather than ‘bottom-up’ angle to the resulting pedagogy. Because of this there is the risk that there may be important features particular to student genres which have not been identified and are thus neglected.

Nevertheless, in their work Paltridge and Bruce raise an important and very real issue in problematizing the treatment of the linguistic level within genre analysis and pedagogy for EAP, especially with regards to student academic writing where there is still relatively limited knowledge regarding how genres are realised lexico-grammatically below the level of ‘moves’, ‘steps’ or ‘stages’. A better solution to this issue, however, is likely to lie not in application of pre-existing linguistic frameworks which have been developed in analysis other types of texts, but in studies involving bottom-up lexico-grammatical analysis of real student texts from different genres and disciplines across the academy.

2.1.5. Changing views on the 'three schools'

With regards to the three main traditions, NR, SFL and ESP, the existence of a clear delineation between these has been increasingly questioned. Hyon's frequently-cited article (1996) is seen as 'seminal' in its differentiating of the three schools or traditions of genre analysis and her delineation continues to be used (Johns, 2008; Bawarshi and Reiff, 2010). However, Johns et al. (2006) have argued that the situation is in fact 'much more complex'; although differences exist, 'particularly in where the experts start (text or context) and what they emphasize in theories and pedagogies', there is in fact 'considerable overlap' (p. 247). Swales (2009, p. 4) argues that by 2007, with the recent publication of books by representatives of all traditions (Bhatia, 2004; Devitt, 2004; Frow, 2006; Swales, 2004), 'what had become known as the genre movement had coalesced somewhat, with the result that the divisions among the three traditions have become much less sharp', and sets out the 'consolidating trends' as follows (p. 5):

- (a) a balance between constraint and choice; (b) the role of local contextual coloring in the realization of genre exemplars, such as the Brazilian predilection for using *Consideracoes Finais* for the final article section title; (c) a greater sense that genres and genre sets are always evolving in response to various exigencies; and (d) a consequent more nuanced approach to genre awareness and genre acquisition.

Important differences nonetheless remain reflecting different priorities in terms of the ultimate purpose in studying genres. The most significant difference in this respect continues to be with regard to the privileging of either context, in the case of NR, or text, in the case of SFL and ESP. It is the latter perspective which underlies the approach taken in this project. In many ways it can be argued that the NR model in which context and social action are emphasised over the product of these, the form, is more compelling as a way of accurately understanding genre; the communicative function leads to the need for a form and not the reverse. However, this is perhaps not taking into account a possible 'form-function cycle' which develops whereby eventually the characteristics of the form are just as

important in influencing and to an extent leading to the nature of the communicative situation or action. In the case of academic genres, because of the relatively lasting nature of academic texts it could be argued that they have a disproportionate influence on other texts that follow, and that academic genre, although evolving, is slowed in this evolution by the existence of previous instances of the genre; awareness of and imitation of the form of previous instances of a genre combine with social action to produce a new instance of a genre. Furthermore, it can be argued that in many soft disciplines the text and language itself to a large extent embodies the action; as Martin argues (1993, p228) grammatical metaphor facilitates the abstraction needed to enable the academic field of history to exist. At a practical pedagogical level for EAP, it seems very probable that, for example, if students of History were to undertake a guided analysis of the structure and lexico-grammatical features of previous instances of texts belonging to the Essay genre produced by strong students from the same discipline at a similar 'level' (e.g. 'third year undergraduate'), they would learn much about the purposes and linguistic realisation of these purposes within the example texts that would be directly applicable to their own writing in the Essay genre within this discipline.

However, there is a broader point to be made about the genre approach, which can get obscured in discussions about the competing traditions within it. Despite genre's ascension in the late eighties, almost three decades ago now, there is a noticeable lag with regards to mainstream EAP's abandonment of 'process' approaches and take-up of genre-informed pedagogies. Tribble's (2015) recent survey of published writing materials provides evidence of this. He notes that whilst Nesi and Gardner's (2012) work is beginning to have an impact, textbooks with a primarily process approach still predominate (2015, p460). For practical purposes textbooks (and this is true of all reviewed within Tribble's article) tend to follow an EGAP rather than ESAP approach and be designed for general pre-sessional contexts. My own experience teaching in-session EAP at two UK HE institutions as well as anecdotal evidence in the form of presentations at BALEAP conferences and SIGs in recent years suggests that at a local, unpublished level a lot more in way of

genre-informed pedagogies exist in instruction of writing on discipline-specific ESAP in-session courses, with materials being created around analysis of a small sample of student texts provided by the target department. However, in order for writing instruction in EGAP contexts to be as effective as possible I would argue that mainstream textbooks and pre-session classes need also to be moving in a genre direction. For this to be made possible similarities and differences between student writing across different disciplines and genres need to be better understood for which Nesi and Gardner's (2012) research represents a significant advance, but more individual studies closely examining particular subsets of disciplines, particularly at a lexico-grammatical level, are needed to further 'flesh out' the picture. In the following section the state of our knowledge of the disciplinary characteristics of academic texts both at a professional and at a student level will be examined in detail.

2.2. Discipline

2.2.1. A model for the disciplines

The model of academic disciplines most widely drawn on in linguistics studies of the subject is that developed by Tony Becher (Becher 1981, Becher 1987a, Becher 1987b, Becher 1989, Becher 1994; Becher and Trowler 2001, Newmann, Parry & Becher, 2002). This model has recently been explicated most fully in Becher and Trowler (2001). Becher and Trowler's (2001) model of disciplinary groupings and disciplinary knowledge provides a useful starting point for investigating the concept of academic discipline. Becher and Trowler (2001, pp. 31-35) criticize the lack of capacity for subtlety of distinction between areas of knowledge afforded by the earlier 'uni-dimensional' nature of models used to describe the sciences developed by Pantin (1968, cited in Becher and Trowler, 2001) who saw a division between 'restricted' and 'unrestricted' fields and Kuhn (1962, cited in Becher and Trowler, 2001) who differentiated paradigmatic from pre-paradigmatic fields. They argue the need for an examination of the disciplines that goes beyond the sciences. They also argue that, in terms of the humanities and social sciences, disciplinary areas that

would be defined by Kuhn's model as 'pre-paradigmatic' are not as he would argue on a path of evolution to 'mature paradigmatic status', but are in fact disciplines in which dissension and pluralism are natural. In this sense one defining feature of a discipline is the extent to which it is inherently prone to consensus or plurality (p. 33).

Becher and Trowler offer a new scheme which synthesises models developed by Biglan (1973, cited in Becher and Trowler, 2001) and Kolb (1981, cited in Becher and Trowler, 2001, pp34-35) both of whom take a perspective of knowledge drawn from research into how those engaged in the activity of a particular arena of knowledge perceive the arena they are engaging with (p. 34), the former drawing on data based on questionnaires conducted with academics, the latter on data regarding the learning styles of students (pp. 34-35). This scheme divides academic knowledge into four broad categories of hard pure, soft pure, hard applied and soft applied fields (p. 36) (as seen in Figure 1.) and in a 'broad-brushed' way delineates the 'epistemological features' of each category in terms of 'characteristics of the objects of enquiry; the nature of knowledge growth; the relationship between the researcher and knowledge; enquiry procedures; extent of truth claims and criteria for making them [and] the results of research' (pp. 35-36).

| Disciplinary grouping | Nature of knowledge |
|---|---|
| Pure sciences (e.g. physics); 'hard-pure' | Cumulative (crystalline/tree-like); atomistic, concerned with universals, quantities, simplification; impersonal; value-free; clear criteria for knowledge verification and obsolescence; consensus over significant questions to address, now and then in the future; results in discovery/explanation |
| Humanities (e.g. history) and pure social sciences (e.g. anthropology): 'soft-pure' | Reiterative; holistic (organic/river-like); concerned with particulars, qualities, complication; personal; value-laden; dispute over criteria for knowledge verification and obsolescence; lack of consensus over significant questions to address; results in understanding/interpretation |
| Technologies (e.g. mechanical engineering, clinical medicine): 'hard-applied' | Purposive; pragmatic (know-how via hard knowledge); concerned with mastery of physical environment; applies heuristic approaches; uses both qualitative and quantitative approaches; criteria for judgement are purposive, functional; results in products/techniques |
| Applied social science (e.g. education, law, social administration): 'soft-applied' | Functional; utilitarian (know-how via soft knowledge); concerned with enhancement of [semi-] professional practice; uses case studies and case law to a large extent; results in protocols/procedures |

Figure 1: Knowledge and disciplinary grouping in Becher and Trowler (2001, p. 36)

Becher and Trowler point out a number of limitations to their framework of the disciplines. They acknowledge that it entails a 'realist' that disciplinary knowledge reflects 'a discernible and stable reality' as opposed to 'phenomenological' view of disciplines as 'essentially socially constructed' (p. 37). Although questioning the relativism of the phenomenological view they do not deny that social factors and contexts, particularly power relations, are important in the shaping of epistemology and argue that in fact social processes mediate knowledge structures (ibid.); as a result of this reasoning they recognise the need to acknowledge not only disciplinary knowledge but also narrative 'stories' regarding disciplinary epistemology (p. 38). They highlight two further caveats in terms of their framework, the first being that the fact the nature of knowledge is continually

evolving means it is difficult to argue that any attempts to classify it can be enduring or permanent, and the second being that the categories and boundaries between these in the framework oversimplify a reality that is in fact much more ambiguous (pp. 37-39). They conclude, nonetheless, that '[t]he Kolb-Biglan framework ... can provide a workmanlike set of categories and a useful basic terminology for exploring knowledge in all its variety ... and ... particularity' (p. 39).

Further to these caveats, they argue that the nature of a discipline is by no means straightforward. The concept has the capacity for uncertainties in its application. These uncertainties can include such matters as when a knowledge area is sufficiently separated from its original disciplinary source to become a discipline in its own right, as in the case of statistics separating from maths, or whether a new disciplinary area, such as for example black studies, is seen as valid (p. 41). Criteria that can be interpreted as evidence to support the existence of a discipline include the existence of professional associations and specialist journals for the area of knowledge seeking the status of 'discipline', whether this area has currency internationally, whether its subject matter is seen as appropriate, and whether it is seen as having 'academic credibility' or 'intellectual substance' (ibid.).

Becher and Trowler acknowledge significant institutional variations in terms of how disciplines are delineated, but reject the argument that this implies the need for a strong form of the 'situationally contingent approach'; they argue that there are more consistent patterns in divisions within departments, such as, for example, between 'pomos' (postmodernists) and 'the rest' in sociology than would be possible if they were completely dependent on their local institutional context (p. 42). They state that at the global level disciplines exhibit both unity and diversity arguing that disciplines vary historically in terms of how knowledge domains change in nature over time, and geographically in terms of 'differences in emphasis' (p. 43). They cite Rusco's (1987) biological analogy of the genotype, 'the fundamental instructions to the organism', and the phenotype, the 'actual manifestation of that potential in a particular physical setting' and his argument that although considerable 'phenotypical variations' exist between disciplines there are

nonetheless 'genotypical' cultural characteristics intrinsic to disciplines. They acknowledge the existence of variations reflecting both features of local educational systems and development levels, and 'national traits and traditions', but point out that '[n]o one who writes or speaks of national differences seems to want to deny that strong resemblances persist between different branches of the same family' (ibid.).

Overall, Becher and Trowler seem to argue that the characteristics that a discipline shares tend to have a stronger influence than potential differences based on institution or geographical location. An important question for the current study is whether this is the case not only at a 'professional' level in academia but also at a student, specifically undergraduate level at UK HE institutions. The strength of cross-institutional continuities in how discipline is instantiated in departments and degree programmes may mean a high level of cross-institutional consistency in the writing of undergraduate students for their degree programmes; conversely, the different nature and status of early student work in a discipline may mean the particularities of local contexts have a stronger influence on student academic writing. Neither of these possibilities can be assumed and therefore need to be taken into account in research that investigates student writing.

2.2.2. Discourse community and disciplinary discourse

Related to any conception of the disciplines is the question of how to conceptualise the individuals working and communicating in relation to them. A useful tool for this purpose is Swales' concept of *discourse community* (1990) (referred to above in discussion of the ESP tradition of genre scholarship). Swales defines discourse communities as 'socio-rhetorical networks' and sees them as possessing 'six defining characteristics' (p. 24): discourse communities have 'a broadly agreed set of common public goals'; they have mechanisms by which their members can communicate with each other; these mechanisms are primarily for provision of 'information and feedback'; a discourse community 'possesses one or more genres

in the communicative furtherance of its aims' and 'has acquired some specific lexis'; lastly, a discourse community 'has a threshold level of members with a suitable degree of relevant content and discorsal expertise' (p. 24-27).

Swales' argument that the activity of those working in a *discourse community* creates distinctive, community-specific linguistic features/communicative patterns links to the concept of *disciplinary discourse* (Becher 1987; Hyland 2000). Becher hypothesizes this link between discipline and language as part of his wider research on the nature of academic disciplines (1987, p. 261):

... one would expect differences in fields of knowledge to be reflected in linguistic form: and by the same token, differences in linguistic form to signify differences in fields of knowledge.

Hyland (2000, p. 8) argues that writing demonstrates legitimate membership of the disciplinary group both tangibly and publicly and involves 'disciplinary constraints on discourse which are both restrictive and authorising ... allowing one to create successful texts which display one's disciplinarity, or tacit knowledge of its expectations' (ibid.). He, however, underlines the need to 'be cautious in emphasising the degree to which consensus exists' (ibid.), instead seeing disciplines 'as systems in which multiple beliefs and practices overlap and intersect' (p. 11) and stating that '[c]ommunities are frequently pluralities of practices and beliefs' (ibid.). This point is also addressed by Bondi who argues that (2006) 'no attempt can be made to map disciplinary variation into a topology with oppositional parameters delineating mutually exclusive categories', and that, rather, '[t]he mapping of disciplines and their discourses' involves 'working out a topology of discourses: sets of criteria for establishing degrees of proximity ... where individual texts can be positioned on a cline, as more or less prototypical'.

Groom points out that although the term 'disciplinary discourse' is the title of both an article by Becher (1987) and a book by Hyland (2000), 'one looks in vain for a working definition of the term' and seeks to remedy this (2007, p. 23). He takes 'discourse' to be (p. 24):

[A]n uncountable noun [which] tends to be used to refer to “any naturally occurring stretch language, spoken or written” (Carter 1995: 39), thereby effectively incorporating all and any linguistic phenomena not covered by mainstream Chomskyan linguistics, with its exclusive focus on clause and sentence grammar Discourse analysis, in this sense, then is “the sociolinguistic analysis of natural language”, as the subtitle of one classic introduction to the field puts it (Stubbs 1983)

This definition is positioning ‘discourse’ at a social level as being language in the process of performing functions and realising purposes attached to particular social contexts. Groom goes on to formulate the following definition for *disciplinary discourse* (p. 25):

[A] stable yet continuing evolving set of meanings, values and practices which produces and is produced by a stable yet continually evolving set of conventional linguistic forms. Disciplinary discourse, in short, is the fusion of epistemology and phraseology.

With this definition Groom can be argued to be positioning language as embodying and producing the social purposes and activities of a discipline.

It needs to be noted that there are alternative conceptions of those working within the disciplines which place less value on the ideas of ‘community’ and ‘discourse’. These approaches can be seen as being in line with the New Rhetoric privileging of context over text in approach to genre. Lave and Wenger’s (1991) term *community of practice*, a sociocultural rather than linguistic term (Borg, 2003, p. 399), places less emphasis on genres and lexis while placing more on shared values and practices. Prior (1994, 1997) rejects the concept of discourse community as too structuralist arguing for the concept of *disciplinarity* which he describes as ‘a sociohistoric approach to writing as literate activity in functional systems and to disciplines as dynamic heterogeneous networks’. He argues (1994, p. 486):

I take the perspective not only that disciplines look more like societies than communities, but also that societies are dynamic historical phenomena ... [and] use disciplinarity rather than discipline to signal an emerging dynamic view of how texts, readers, and writers constitute and are constituted by social formations.

However, the broad consensus of many whose research involves contrastive linguistic analysis of professional academic writing is that the concept of 'discipline' and emphasis on 'discourse' are both useful. Hyland argues (2000, p. 10):

Irrespective of whether we choose to label disciplines as tribes, cultures, discourse communities or communities of practice, these concepts move us from a concern with the abstract logicity and substance of ideas of academic writing to a world of concrete practices and social beliefs. They put community decision-making and engagement at centre-stage and underline the fact that disciplinary discourse involves language users in constructing and displaying their roles and identities as members of social groups.

Bondi (2006, p. 49) argues that from the perspective of those working within the 'intrinsically applied' framework of EAP 'interest lies in way academic and professional communities define their areas and their conventions'. Similarly, Groom (2007, p. 16) argues that whilst acknowledging alternative perspectives 'we will nevertheless continue to regard the concept of discipline as a valid and useful one, at least for research projects (such as this thesis) whose focus is at the macro-social rather than micro-social or ethnographic level'.

2.2.3. Disciplinary variation in professional academic writing

Charles Bazerman (1981) and Tony Becher (1987) provide two highly influential early studies of the link between discipline and variations in discourse. In his 1981 article, 'What Written Knowledge Does: Three Examples of Academic Discourse', Bazerman closely analyses three journal articles from three respective disciplines, molecular biology, sociology and literary criticism, taking a bottom-up approach to identifying variation based on 'the object of study, the literature of the field, the anticipated audience, and the author's own self' (1981, p. 362). He concludes that '[i]n mediating reality, literature, audience and self, each text seems to be making a different kind of move in a different kind of game' (p. 378). In terms of the 'reality' being dealt with, in the molecular biology article the independent existence of the phenomena in focus, DNA and genetic carrier, is collectively accepted from the

outset of the article, whereas in the sociology article focusing on the concept of 'ambivalence', the author 'must first establish that the phenomenon exists and is consequential', and in the literature article focused on a Wordsworth poem, 'what was known should not count as true knowledge, which can only come in the subjective recreation of the poetic moment' (Bazerman, p378).

In terms of field, audience and the author identity, in the molecular biology article there is no need for explicit discussion of the greater part of the relevant literature, with only 'claims and evidence immediately bearing on the essay's claim' needing to be attended to (ibid.), and it is written to an audience who jointly accept the same body of knowledge, techniques for evidence gathering and judgement criteria (ibid.); thus, the authors assume the identity of 'humble servants of nature and their discipline', making a small contribution to the collective building of a vast picture being collectively whilst 'subject to the hard evidence of nature and the cold judgement of their peers' (Bazerman, p378). In contrast, the sociology article is written in a field where the literature is 'more diverse, unsettled, and open to interpretation' and, therefore, for a discussion framework to be established the literature must be reconstructed. It is written to an audience which does not share a uniform thought framework or proof criteria. They therefore 'must be urged, persuaded and directed along the lines of the author's thoughts'. Thus, in terms of disciplinary identity, the author 'stands more uncertainly before his discipline and nature, neither of which holds the promise of clear-cut judgement or unequivocal support' (Bazerman, p378). The literature article, in contrast again, in drawing on poetry and criticism of poetry, does so 'idiosyncratically and only in support of the critics vision of the particular poetic moment of consciousness being investigated' (ibid.). Therefore, in terms of identity, Bazerman argues the author of the literature article has to assume 'the most demanding role' in conveying the impression of possession of greater insight than his audience: '[s]ince his contribution cannot be measured in terms of a claim to be judged right or wrong, the quality of his whole sensibility is up for judgement' (pp. 138-139). Bazerman concludes (p. 139):

The diversity of knowledge-producing activity embodied in these three texts suggests how important the form of knowledge is. Getting the words right is

more than fine tuning of grace and clarity; it is defining the entire enterprise. And getting the words right depends not on an individual's choice. The words are shaped by the discipline – in its communally developed linguistic resources and expectations ... The words arise out of the activities, procedures, and relationships within the community.

Bazerman's study takes a rhetorical rather than linguistic approach so does not attempt to link disciplinary differences with any particular recurring linguistic features within the focal texts.

Becher's (1987) study seeks to test his hypothesis (outlined above) about disciplinary differences and the link between these and discourse. He interviews scholars from the fields of physics, sociology and history about how they judge the quality of academic work within their fields and then compares their answers with what is written in book reviews taken from leading disciplinary journals in each field. He identifies differences in the adjectives deployed in the respective fields to praise or criticise work (1987a, p. 263). For example, 'rigorous' has positive connotations for historians, whereas for physicists the same word is a 'backhanded compliment' which implies 'a narrow, unimaginative employment of mathematics with no proper grounding in the physical world'. Although historians and sociologists share use of some positive adjectives, such as 'rigorous' and 'stimulating', Becher argues other evaluative adjectives favoured only by sociologists such as 'persuasive' and 'thought-provoking' are suggestive of especial interest in the analysis in and of itself or its impact on the audience as opposed to 'its substantive content' (ibid.). Becher argues that how the vocabularies of these disciplines contrast provide an indication of cultural differences between them (ibid.).

Becher also analyses the leading journals from each discipline and concludes similarities at surface level fail to hide important disciplinary differences with regard to presentation and structure (p. 266). He notes, for example, differences in the length of articles and frequency of publication, finding physics articles noticeably shorter and published in much higher volume and frequency, with some journals producing 500 to 600 fortnightly compared to the relatively slower and more

modest quarterly output of leading history and sociology journals (p. 267). In terms of text content he notes differences in language, 'impenetrable to the outsider' in physics, 'easily accessible – at least on the surface' in history, and 'less readily intelligible to the layman' due to deployment of technical and semi-technical terminology in sociology (p. 268). He finds writers in the three disciplines use similar levels of referencing, but in ways that differ systematically which, again, underline disciplinary differences: 'the expectation in history of close patterning on the primary data; the volatile nature of the sociological argument ...; and the tight mesh of research in physics, mobbing almost inexorably from one set of findings to another' (pp. 268-269). His conclusion supports his hypothesis of the link between the disciplines and their discourses (p. 273):

In summary, the three disciplines display fundamental differences not only between types of evidence and procedures for proof, but also in the ways in which others' work is evaluated and in the modes in which arguments are generated, developed, expressed and reported. All these have implications for the underlying knowledge structure, indicating whether it is dense or diffused, atomistic or holistic, stable or volatile, universal or particular.

However, Becher's conclusions about the disciplinary orientation of linguistic features and referencing purposes within journal articles are a combination of his own observations and anecdotal evidence drawn from comments of his interview subjects; they are not the result of any form of systematic linguistic analysis of the texts.

Linguistic analysis of professional academic writing with regards to disciplinary features can take a broadly Swalesean approach to interrogating schematic structure, or an approach which involves closer analysis of sentence-level lexicogrammatical features. Examples of Swalesian-influenced work include that by Samraj and Basturkmen. Samraj (2002) investigates the introductions of research articles at the level of macro-structure in two related fields, that of Wildlife Behaviour and Conservation Biology, using Swales' Create-A-Research-Space (CARS) model, and finds disciplinary variation at this level of analysis. Basturkmen (2012) also works at the macro-structural level, comparing discussion sections in research

articles in Dentistry and Applied Linguistics, and also uses a Swalesean 'moves and steps' framework for analysis finding a similar pattern of moves across the two disciplines, but differences at the level of steps and sub-steps (p. 143). In related work using his own *social genre/cognitive genre* model to compare the results sections of sociology and organic chemistry articles, Bruce (2009) finds that these two disciplines differ significantly with regard to the 'social genre elements of context, epistemology and writer stance' and, in terms of cognitive genre, finds the two disciplines differ consistently with regard to employment of textual resources with chemistry results sections mainly employing the *explanation* cognitive genre and sociology results having a preference for the *report* cognitive genre (p. 105).

The most well-known and large-scale piece of research into professional writing at the lexico-grammatical level is Hyland's work (e.g. 2000, 2001, 2002, 2005, 2007, 2008, 2009). Taking a corpus-based approach Hyland identifies differences between professional writing across the sciences, social sciences, and humanities at the word and sentence level of the text. In a corpus analysis of 1.5 million words taken from research articles, book reviews, scientific letters, abstracts, and textbook chapters from the disciplines of Molecular biology and magnetic physics (representing the pure sciences), Mechanical Engineering and Electronic Engineering (representing the applied sciences), Philosophy and Sociology (representing humanities and social sciences), and Marketing and Applied Linguistics (representing the more applied social sciences) (page xi) he examines disciplinary difference including focus on 'citation, reporting verbs, hedges, self-mention, directives, and lexical bundles' (2009, p. 10).

Hyland finds a much greater proportion of citation, two thirds of instances of this within the entire article corpus, in the 'soft' disciplines of humanities, social sciences and applied social sciences (2009) which he argues reflects differing extents to which it can be assumed that a context is shared with readers; literature in the soft disciplines is dispersed with a more heterogeneous readership meaning a shared context cannot be presupposed but has to be built through citation to a far greater extent. He also finds that the verbs used to refer to the literature differ significantly

on disciplinary grounds and appear to reflect the broad purposes of disciplines with verbs typically used in the soft fields, such as *discuss*, *hypothesize*, *suggest* and *argue* reflecting 'writing activities', while those used in the hard disciplines, such as *observe*, *discover*, *show*, *analyse* and *calculate*, emphasizing real-world activities and thus 'representing knowledge as proceeding from impersonal lab activities rather than from interpretations of researchers' (pp. 11-12). Hyland finds further disciplinary divergence in the frequency of use of hedges such as *possible*, *might*, and *likely* which occur twice as frequently in humanities and social science writing compared to its frequency hard sciences writing. He attributes this difference to the fact that in the soft fields variables are harder to control, research outcomes are more diverse and there are 'fewer clear bases for accepting claims' meaning in the soft fields the need to express arguments more cautiously than in the hard fields (p. 13), and also to the fact that because of the positivist epistemologies of the hard sciences individual authority is in a subordinate position to the authority of the text and it is meant that facts 'speak for themselves' meaning the interpretive activities of writers are disguised behind linguistic objectivity (p. 13).

Hyland also finds preferences that are broadly disciplinary in levels of self-mention; as with citation, two-thirds of the instances of self-mention occur in the soft disciplines (p. 14). He attributes this difference to the far greater dependency in the soft fields on being able to 'invoke the sense of a real writer in the text' and present 'a confident and authoritative authorial self' (p. 15). This contrasts with the way in which those working in the hard sciences 'downplay their personal role in the research to highlight the phenomena under study, the replicability of research activities, and the generality of findings' (pp. 15-16). Another area of difference is in the use of 'directives', devices used to instruct the reader to either view things in a certain way or perform certain actions: most directives in the soft fields are 'textual', such as *see* and *refer*, whereas there are a high proportion of cognitive directives, such as *note*, *concede*, and *consider*, in the hard fields (pp. 17-18). He explains this difference as relating to writers in the soft disciplines being 'less able to rely on the explanatory value of accepted procedures' than those in the hard fields where argument 'is formulated in a highly standardised code' (ibid.); the

writer instructing readers to see things in a particular way could be seen as 'assuming unwarranted authority' in the soft disciplines, whereas the importance of succinctness in the hard disciplines make cognitive directives useful for cutting 'directly to the heart of the key issues in the text' (Hyland, p18). There is also a high level of disciplinary specificity in terms of use of lexical bundles such as *on the other hand* and *in the case of* which Hyland argues are key to the shaping of meaning in a text and to contributing to a sense that the register is distinctive and natural. More than fifty percent of the top fifty bundles in each of the disciplines are not found in the top fifty bundles occurring in other disciplines, and the greatest similarities occurring between soft disciplines and hard disciplines respectively (ibid.).

In terms of Becher's disciplinary framework, Hyland's study provides a broad survey of discipline that spans the 'hard' to 'soft' spectrum. However, it is important to remember his study nonetheless involved eight discrete disciplines, so the qualities that are attributed in a wider sense to the 'hard' or 'soft' nature of each need to be treated with a certain degree of caution. Other studies of disciplinary features of professional academic writing at the lexico-grammatical level typically involve a smaller range of disciplines. Many of these studies involve contrasting disciplines situated in different parts of the disciplinary matrix (e.g. Conrad, 2001; Cortes, 2004; Charles, 2006; Silver, 2006). For example, Charles (2006) investigates *that*-clause complementation for reporting (p. 492) in the social science, politics and the natural science, materials science, finding a significantly higher use of 'human' subjects in politics and a higher use of *it* subjects in materials science (p. 498). Silver (2006) investigates two 'hard' disciplines, unified physics and molecular biology, and two 'soft' disciplines, business and economics with the intention behind the inclusion of two from each area being to take in account 'meta-disciplinary' factors. He asserts a need to contest the widely-accepted idea that there will be more methodological and argumentative similarities in disciplines sharing a meta-disciplinary area than there will be in disciplines across meta-disciplinary areas. Interestingly, in line with his hypothesis, he finds greater similarities in terms of textual features between Physics and Economics than Economics and Business; Both Physics and Economics use a high number of 'meta-cognitive verbs signalling

prediction' whereas they are all but absent from Business and Molecular Biology (p. 98). He attributes this finding to the difference between the hypothetical world of Physics and Economics and the empirical world of Business and Molecular Biology (ibid.).

Another set of studies involve the contrastive analysis of disciplines that sit relatively closely on the disciplinary spectrum (e.g. Bondi 2004, 2006; Bondi and Silver, 2004; Groom, 2005, 2007; Nishina, 2010; Malavassi and Mazzi (2010). Bondi (2006, p. 52) investigates specificity in the academic discourse of business in contrast to that of economics using 'smaller corpora' to investigate generic structure and 'larger corpora' to quantitatively investigate lexical units in context, which is what will be discussed here. She conducts a key-word analysis of the Business and Economics corpora and focuses in on grammatical items from the lists generated. She attributes the high placing of *were* on the Business key-word list to the discourse community's 'typical discursive procedures' most likely related to deployment in narrative case studies (p. 64) and argues the high frequency of *if* and *is* in Economics is '[d]ecisive proof of the speculative nature of economics' (p. 66). However, she does not undertake close investigation of concordances of these grammatical words, so her conclusions are hypothesised rather than proved. Bondi does undertake close analysis of concordances of lexical items pre-determined for their role as 'labelling nouns referring to narrative sections of text' (p. 55) such as *case* and *history* and finds significant differences both in frequency of particular items and semantic patterning around items concluding that variations across disciplines is observable even in analysis of such close disciplinary neighbours as business and economics (although, as seen in the previous paragraph, Silver (2006) argues that these two disciplines' respective empirical and hypothetical orientations means differences are quite understandable).

Both Groom (2007) and Malavasi and Mazzi (2010) use keywords as a starting point to investigate differences between two close disciplines. Groom (2007) undertakes keyness analyses for two large disciplinary corpora composed of research articles from the disciplines of History and Literary Criticism. In a qualitative concordance

line analysis of grammatical items key in each corpus he identifies the different types of semantic patternings occurring for each item and both similarities and differences between the disciplines in terms of preference for particular meanings. Groom makes links between the phraseologies shared across the two disciplines and the *reiterative*, *holistic* and *particularistic* characteristics that Becher (1989) identifies as embodying the epistemology of the 'soft-pure' humanities domain to argue that there are 'pervasive, clear and consistent relationships between the phraseology and epistemology of the disciplinary discourses of history and literary criticism' (p. 282) (N.B. this study will be discussed in considerably more depth in the Methodology chapter as the approach involved has been adapted for my project).

In contrast to Groom's focus on grammatical words, Malavasi and Mazzi (2010) focus on key lexical items in corpora of History and Marketing research articles which they see to be embodying important content meaning. They select five items from the keyword lists for each discipline, *he*, *science*, *historians*, *text* and *society* for History and *we*, *results*, *effects*, *research* and *data* for Marketing (p. 173) and investigate collocational patterns for these finding differences that are disciplinary in 'agentive subjects, objects and research procedures' (p. 167).

Taking a corpus-based approach Nishina (2010) also compares professional writing across two relatively similar disciplines, this time the 'soft-applied' fields of applied linguistics and business studies. He analyses the pattern of adjectives followed by prepositions and the collocations within and around them. He finds disciplinary differences in terms of both form and meaning which he argues are indicative of the distinct epistemological characteristics and cultures and norms of the two disciplines.

This section has provided a survey of research into the disciplinary characteristics of professional academic writing including the important early studies of Bazerman (1981) and Becher (1987), the influential corpus work done by Hyland (e.g. 2000, 2001, 2001, 2005, 2007, 2008, 2009), and a range of other studies making use of

corpora that have followed. Given the volume of research in this area it is by no means a comprehensive survey, but has nonetheless sought to provide as accurate picture as possible of the type of research thus far conducted both at the level of genre moves and steps and at the lexico-grammatical level. All studies discussed above have involved the genre of the RA. This is the key professional academic writing genre, the central form of communication for the discourse community of a discipline. Researchers have thus been able at the outset to make an assumption that what they are dealing with is 'disciplinary discourse'. It would, however, be dangerous for researchers of student academic writing, particularly at undergraduate level, to make any similar assumption. In the following section the unique position and nature of student as opposed to professional academic writing is discussed before a survey is provided of studies to date in this relatively much more nascent area of research.

2.2.4 Student writing, community and discipline

'Student writing' is a broad category taking in as it does first year undergraduate writing to PhD writing. For this reason, level is extremely important in indicating how similar or different in nature student texts are to professional, published writing in the same disciplines. Master's theses were excluded from the BAWE corpus partly because they 'often underwent a process of drafting, redrafting and review which made them more like published texts' (Alsop and Nesi, 2009, p. 74). For the purposes of this project which investigates third-year undergraduate texts, the review of literature will focus particularly on issues regarding and findings about writing at undergraduate level.

Arguably, student texts are written within a 'disciplinary' context delineated both by the degree programme the student author is enrolled on and the academic department hosting this degree programme. However, the relationship between student texts and notions of disciplinary community and discourse discussed above is an inherently problematic one. Both Johns (2008) and Nesi and Gardner (2012)

make important points about how the 'student' status of texts affects their relationship to the discipline they are ostensibly written within. Johns (2008, p. 240) argues that the low status of student texts within academia leads to careless genre labelling:

Unfortunately, student-produced genres in academic contexts are much more casually named by their instructors than are respected academic genres, probably because student texts have little or no prestige in academic communities ... instructors across the disciplines call may examination and out-of-class papers 'essays', when, in fact, the appropriate structure, register and argumentation in these papers will vary across classes and disciplines. What is an essay? This is a very difficult question for us to answer; and because student essays do not really matter to disciplinary experts, they do not consider the question.

Nesi and Gardner (2012, p. 23) argue that the 'broad social purpose', that of accreditation, unique to student assignment writing, differentiates it from other disciplinary writing such as that of textbooks and research articles. They also describe student assignments as 'high-stakes texts' because failing in an assignment can lead to failing in a degree entailing costs academically, socially and financially (ibid.). This 'low prestige' but 'high stakes' status very clearly demarcates student texts from professional activity within a discipline.

Another problematic idea with respect to student academic texts, particularly undergraduate texts, is the notion that by undertaking the writing required on their degree programme students are participating in an 'apprenticeship' into the discourse community of their discipline through the process of which they progressively acquire its discursive practices. Berkenkotter and Huckin (1995, p. 13) argue the acquisition of disciplinary genres requires 'immersion into the culture and a lengthy period of apprenticeship and enculturation' which they term 'cognitive apprenticeship'; however, they see this model as applying to graduate rather than undergraduate students due to the fact the latter learn through pedagogic rather than completely authentic disciplinary genres. This argument about apprenticeship is supported by Candlin and Plum's (1999, cited in Woodward-Kron, 2004, p. 142)

findings regarding students at undergraduate level in Psychology; these students did not generally feel they were being 'apprenticed' in Psychology as a discipline lacking opportunity for genuine participation in a peripheral sense in the discipline. Woodward-Kron's (2004) study investigated the notions of 'discourse community' and 'writing apprenticeship' in relation to undergraduate students of Education at an Australian university using data including marker feedback on assignments over the three years of their degree and interviews with both tutors and students. She found that although many of the comments made by markers on assignments 'had a socialising dimension' (p. 151) the amount of this kind of feedback was actually small relatively (p. 152). Interviews with tutors revealed that they largely saw themselves as socialising their students into their future professional role as teachers rather than into the academic discipline of Education; for example, the function of a third-year task involving the review of a research article was seen as training students to be informed consumers of academic research rather than potential future authors of such texts. Similarly, in interviews with students, most saw this same research article task as preparing them for the real-world classroom situations. Interestingly, however, the highest performing student in the study saw this task differently as 'a definite push' to encourage students to see themselves as both reflective practitioners and researchers and to think about progressing to postgraduate level (p. 156). This suggests that whether or not an 'apprenticeship' is taking place at undergraduate level could possibly be argued to be very much dependent on individual student motivation and self-identity; a student considering a continuing pathway in the discipline may see the purpose of his/her undergraduate writing very differently a student in the same programme with a more instrumental, short-term personal relationship to the same field.

Russell (1997) sees identity issues as of central importance with respect to how a student approaches a discipline and the writing expectations within it. There may be conflict between the social identity a student carries with them into the new environment of the university and the identity they would need to adopt to operate successfully whilst there (p. 532):

Students may be forced to choose a direction away from activity systems of family, neighbourhood, and friends that construct ethnic, racial, gender and class identity(ies). This forced choice may cause students to feel double binds. What is considered learning to the university or disciplinary activity system may be perceived by other activity systems of family or neighbourhood as selling out.

He argues (pp. 533-534):

The activity theory of genre I am developing suggests that individual students feel the contradiction between and among activity systems of school and society as double binds about whether to involve themselves primarily as consumers of a discipline's or profession's commodified tools (knowledge) in distant genres, to place themselves on the periphery of its activity system, or to become involved actively in its life through deeper participation – to throw *themse/ves* [Russell's italics] into it through the reading/writing of its genres, to make a difference as well as make a grade.

The implications for a student's wider social identity of engaging with a discipline and in doing so complying with the constraints of its writing is also addressed within academic literacies research which focuses particularly on the experience of non-traditional students within the UK Higher Education context (e.g. Lillis, 1997; Ivanic, 1998). Woodward-Kron, however, points out that this same sense of 'loss' of identity can also be experienced by what would be considered 'traditional', non-marginalised undergraduate students (2004, p. 157).

Another way in which academic literacies research is useful to understanding how student texts relate to ideas of 'discipline' is its focus on the complexities involved in untangling and evaluating the relative influences of discipline, institution, department and individual course or tutor on the development of a student's academic writing. This can be seen in the work of Lea and Street (1998). They discuss the powerful impact of implicit disciplinary assumptions on how a piece of student writing is evaluated; they offer the example of a student writing first-year essays for both History and Anthropology whose writing for History was received positively, but whose writing for Anthropology was deemed to lack structure or argument and 'pathologised' as to do with the students' lack of essay-writing skills

for which remedial help was suggested (1998, pp. 165-166). Lea and Street, however, argue that 'the explanation for this divergence of opinion and response lies at a deeper level than the surface features of 'writing' to which the anthropology tutor refers' (p. 165); despite the tutor's comments being 'couched in terms of writing problems', his/her reason for negatively appraising the student's writing are in fact to do with an 'epistemological supposition regarding academic disciplines' which remains concealed behind technical focus on the allegedly generic characteristics of 'academic writing' as a whole across disciplines (p. 167). However, Lea and Street also discuss the powerful influence of the institution on the nature of student writing and how this possibly mediates disciplinary influence (p. 170):

[w]e consider the analysis of writing in the university as an 'institutional' issue ... The institution within which tutors and students write defines the conventions and boundaries of their writing practices, through its procedures and regulations (definitions of plagiarism, requirements of modularity and assessment procedures etc.), whatever individual tutors and students may believe themselves to be as writers, and whatever autonomy and distinctiveness their discipline may assert.

With respect to an examination of the nature and extent of guidance contained in departmental writing guidelines documentation and instructions for particular writing tasks, Lea and Street (1999) consider as well the influence of the idiosyncrasies of individual departments and tutors within institutions. The influence of individual disciplinary tutors is also discussed by Nesi and Gardner (2012, pp. 32-33) who argue that:

lecturers from the same department value different, and in some cases, contradictory features in academic writing – the role of signposting is a typical example where some feel strongly that this is a positive feature and others prefer texts with very little metatext.

In summary, when investigating undergraduate student writing as opposed to professional academic writing the idea that writing is 'disciplinary' or that texts are produced 'within a discipline' is a much more complicated and contested one. In order to proceed with investigations of student writing which involve evaluating the extent to which discipline accounts for particular characteristics of student texts, a

researcher should qualify the use of 'discipline' as being constrained by the limited extent to which any undergraduate text, because of its differentiated status and the complexity of influences that come to bear upon it, can be argued to be a participant in disciplinary activity. In other words, a student text can only be 'disciplinary' within the specific internal logic of student writing. Additionally, investigations of student writing should not assume 'discipline' to be the sole or most influential factor in accounting for particular text characteristics; research design should, if possible, take into account other non-disciplinary dimensions of potential influence.

2.2.5. Research into disciplinary variation in undergraduate student writing

Until relatively recently most research investigating student written assignment tasks and genres across disciplines drew on data such as surveys of students (e.g. Kroll, 1979), interviews with teaching staff (e.g. Braine, 1989), and analysis of departmental documentation outlining task requirements (e.g. Braine 1989; Hale et al., 1996; Moore and Morton, 1999, 2005) rather than examination of the actual student-produced texts written in response to these tasks. For example, Hale et al. (1996) carried out a large scale survey of tasks across courses and universities in the US with the purpose of informing the redevelopment of the TOEFL test, and Moore and Morton (1999, 2005) carried out a survey of disciplinary tasks on a similar scale across two Australian universities as part of a comparison of the requirements of university tasks in comparison to requirements of Task 2 in the academic version of the IELTS test written component. Carter's (2007) work on meta-genres can also be grouped in this body of research which draws conclusions from data other than the actual student texts produced. Working within the New Rhetoric tradition of genre research, he draws on data produced as the outcome of an initiative in which Writing in the Disciplines (WID) professionals worked with departments at an American institution to instigate outcomes-based assessment which involved description and measurement of the demonstrable skills and knowledge expected of students by the end of their degree programmes (p. 387). Using this data he

classifies texts within his particular higher educational context in the United States as belonging to four 'metagenres' reflecting the 'metadisciplines' of his institution. These four metagenres are 'responses to academic learning situations' that respectively require 'problem solving, empirical enquiry, research from sources, and performance'.

The most comprehensive analysis of actual student-produced texts at undergraduate and taught postgraduate level has been undertaken by Nesi and Gardner (2012) as a part of the British Academic Written English (BAWE) corpus project (the contents of the BAWE corpus will be discussed in detail in the Methodology section of this thesis below). Nesi and Gardner's concern at the level of whole texts in their study of BAWE is primarily one of identifying genres across the disciplinary spectrum (p. 29):

As our aim was to understand the nature of assessed writing across the academy, our classification proposes genre families at a level of delicacy that groups similar genres across disciplines.

They identify and delineate the range of genres written by British university students across sixteen disciplinary areas. As discussed previously, Nesi and Gardner adapt the SFL genre model and also draw on ESP genre and academic literacies research in creating a typography of thirteen genre families (p. 26) grouped under five broad social purposes within the student academic context, demonstrating knowledge and understanding, developing powers of informed and independent reasoning, developing research skills, preparing for professional practice and writing for oneself and others (p. 27). In terms of disciplinary variation, Nesi and Gardner identify similarities and differences both in the genres favoured by particular disciplines and in the range of genres employed within a particular discipline. They find, for example, that the majority of History and Sociology assignments belong to 'the Essay genre family' in contrast to a more diverse range of assignments in Engineering entailing genres from across the thirteen genre families (p. 29). An assumption made in their study that should be borne in mind is that findings about the characteristics of texts in each disciplinary sub-group within BAWE are generalizable across British higher education institutions. Nesi and Gardner describe

their book as ‘an overview of the kind of writing British students produce’ (p. 2); however, corpus holdings for each discipline are, where possible, taken from only one of the four institutions in which texts were collected (p. 7).

As part of their examination of particular genre families, Nesi and Gardner discuss both distinction between sub-genres and disciplinary variation. For the purposes of this literature review their discussion of the Essay genre family will be used as an example of this level of their analysis. The essay genre family is a useful focus due to the fact it is the primary genre of student writing in PIR and History, the two disciplines that are the focus of this project. Nesi and Gardner discuss the distribution of the Essay genre across the sixteen disciplines, finding that there are noticeably more instances in the Arts and Humanities and the Social Sciences (601 and 444 texts respectively) than in the Life Sciences or Physical Sciences (127 and 65 instances respectively) (p. 95). After describing the overall three-part structure of this genre family, entailing an introduction, a series of arguments, and a conclusion, they identify six types of essay within the family ‘drawing heavily on the work of Coffin, (2004, 2006) and Martin (1992)’ (p. 98) differentiating each by both approach and stages entailed (pp. 97-100). These are labelled exposition, discussion, challenge, factorial, consequential and commentary. The ‘exposition’ genre, for example, ‘sets out the claim or thesis it argues for in the introduction’ (p. 99), while the ‘discussion’ genre ‘involves entertaining alternative positions’ (p. 100). Nesi and Gardner concede that there is some overlap and ambiguity between such categories so that for some questions either an ‘exposition’ or a ‘discussion’ approach could be taken, and that some ‘challenge’ questions could also be answered by the exposition genre. Examples from a range of disciplinary groupings are given in discussion of each essay sub-genre, but there is no clear indication of how the six essay sub-genres are distributed across disciplinary groupings. Nesi and Gardner draw on the SFL concepts of the hyperThemes and hyperNews (Martin and Rose, 2003, cited in Nesi and Gardner, p. 109) to explain two important features of paragraphs in the Essay genre family, the initial sentence, the hyperTheme, often described as a ‘topic sentence’, which introduces a new topic and claim which is then followed by evidence, and the hyperNEW which concludes the paragraph with

a more explicit claim or argument regarding the contribution of this evidence to the development of the essay argument (ibid.).

Nesi and Gardner end their analysis of the Essay genre family with further discussion of disciplinary difference, stating that they are assuming that differences with regards to what 'evaluation' and 'argument' are in different disciplines will be reflected linguistically in student texts. They outline differences with regards to the use of headings: essays make up 11% of Physical Sciences texts, but 58% of these texts use headings, whereas in the Arts and Humanities where 82% of texts are essays, only 14% have headings (p. 112). They suggest this greater use of headings in Physical Sciences essays could be explained by the influence from other genres used more frequently by Physical Sciences students such as reports (ibid.). Eleven disciplinary groups are compared in terms of the frequency of use of the first person pronoun 'I'; Philosophy has the highest instance, 34.8 per 10,000 words and biological Sciences the lowest with 2.6 per 10,000 words, and the overall trend is that of a higher occurrence of 'I' in Humanities than Sciences (p. 114). Nesi and Gardner point out that '[t]his strong authorial presence guiding the reader through the argument is consistent with earlier studies of master's dissertations (Samraj, 2008) and research articles (Chang and Swales, 1999; Hyland, 2002)'. They compare reasoning across a selection disciplinary groupings through analysis of 'if ... then' sequences arguing 'the nature of the argument from evidence to claim is different ... in each case', 'logical deduction' in Philosophy, 'legal consequence' in Law, 'persuasive interpretation' in English and 'hypothetical inferencing' in Archaeology (p. 118). They conduct a keyword analysis, using the BNC as reference corpus, across the disciplinary grouping in which the Essay genre is popular, creating lists of key verbs, adverbs, adjectives and nouns (p. 126) and infer disciplinary differences from these lists. For example, with respect to adverbs, they argue (p. 127):

It seems that argumentation in Law is distinct in its focus on what is *consequentially, legally right*; in Philosophy on what is *absolutely, morally or logically accepted*; and in Sociology on what is *arguably, fundamentally, increasingly, importantly, socially and predominantly challenged*.

In comparison to their treatment of genres, Nesi and Gardner's analysis of disciplinary differences in the instantiations of these is much less comprehensive in its coverage. Analysis at this level is selective, with a somewhat arbitrary focus on 'I', and also somewhat limited in depth, with the conclusions drawn from disciplinary keyword analyses being speculative rather than supported by a detailed examination of concordance lines. Analysis at this level seems to serve possibly more as an example of the potential types of further analysis that could be conducted with the corpora. In separate work on sub-corpora of BAWE Gardner (2008, 2012) conducts more detailed SFL-style contrastive disciplinary analyses. She conducts a contrastive 'Angle on Field' (Martin, 1993) analysis of sentence subjects of undergraduate student writing also differentiated by year, finding that her results reveal 'disciplinary differences and similarities of the type already documented' (p. 16) in similar studies of professional academic writing (McDonald, 1994; Gosden, 1993; Lewin, Fine and Young, 2001, cited in Gardner, 2008). Gardner (2012) also investigates how the SFL conception of register can illuminate disciplinary variation in an analysis of short extracts from the methods sections of experimental reports in the disciplines of Psychology, Applied Linguistics and Chemistry. She identifies within the extracts the ideational, interpersonal or textual function of particular words and phrases which relate to the three elements of context of situation, the Field, Tenor and Mode (p. 59). For example, in terms of interpersonal features reflecting Tenor, only the Applied Linguistics text used the personal pronoun *I* and the modal *would*, and there were disciplinary differences in terms of the type of appraisal resources employed with Applied Linguistics using affect, *interesting*, Psychology using judgement, *deviant*, and Chemistry using measurement, *four sets of*. In terms of textual differences reflecting Mode:

The denser language of Psychology reflected the greater use of nominalisation, passive voice and embedding when compared with the less dense [Applied Linguistics text] with its personal subjects and active voice, and with the less syntactically complex [Chemistry text] with its alternation of full clauses and items listed as nominal groups (e.g. Tube 1: ...). [The Chemistry text] brings us closer to the material setting and physical events of the experiment; while [the Applied Linguistics text] brings us closer to the writer as an active participant.

The most comprehensive linguistic analysis conducted as part of Nesi and Gardner's (2012) joint project and book is a Multidimensional (MD) analysis based on Biber's original dimensions (1988, cited in Nesi and Gardner, p. 11), *Involved, Narrative, Elaborated, Persuasive, and Abstract and Impersonal*, which was conducted at the level of the four disciplinary groupings of the BAWE corpus, Life Sciences, Physical Sciences, Arts and Humanities and Social Sciences (pp. 11-13). In terms of disciplinary variation, they find that (p. 14):

texts in the Life Sciences (LS) are the most informational (that is, least involved), and those in the Arts and Humanities (AH) have the greatest amount of narrative features. Physical Sciences have the fewest narrative features and are most impersonal and persuasive. Texts in the Social Sciences (SS) are the most elaborated.

The fact that this analysis was not also conducted at the level of individual disciplines means it cannot offer insights at the finer levels of delicacy that many studies of disciplinary characteristics in professional academic writing have shown such as potential the differences between 'close' disciplines (e.g. Groom, 2007; Nishina, 2010) or potential similarities between disciplines from different groupings (e.g. Silver, 2006).

Hardy and Römer (2013) have undertaken a Biber-influenced dimensional analysis of student writing at the level of specific academic disciplines. In contrast to Nesi and Gardner's (2012) use of dimensional analysis, however, they do not adopt Biber's original dimensions in unmodified form, but follow Biber's process to create their own set of dimensions within student writing. Hardy and Römer (2013) identify four dimensions in student writing collected for the Michigan Corpus of Upper-Level Student Papers (MICUSP), which consists of fourth-year undergraduate and postgraduate coursework writing in a US context; these four functional dimensions 'appear to distinguish between (1) Involved, Academic Narrative *versus* Descriptive, Informational Discourse; (2) Expression of Opinions and Mental Processes; (3) Situation-Dependent, Non-Procedural Evaluation *versus* Procedural Discourse; and (4) Production of Possibility Statement and Argumentation' (p. 183).

Each dimension either positively or negatively associates with particular combination of grammatical features; for example, Dimension 4 positively associates with features including uninflected present, imperative and third person verbs, modals of prediction and predicative adjectives, and negatively associates with past tense verbs.

The sixteen disciplinary groupings in MICUSP were scored and placed on a positive to negative cline for each Dimension and this process revealed significant disciplinary variations. However, whilst in Dimension 3 there is a reasonably clear distinction between the humanities and the sciences (p. 197), for the other Dimensions this distinction was more muted with many exceptions or overlaps in the 'hard' and 'soft' distinction and with the differences being a matter of degree rather than absolute. For example, for Dimension 1 Philosophy and Education score a very positive 8, while a cluster of disciplines from across the spectrum, Linguistics, Industrial and Operations Engineering, Sociology, Nursing and English, cluster between -2 and 2, History and Classical Studies scores -3, Economics -4, Biology -6 and Physics -8 (p. 192). Hardy and Römer's results highlight the dangers inherent in conducting such a multi-dimensional analysis at the level of pre-assumed meta-disciplinary groupings as Nesi and Gardner have done (2012).

Usefully, the BAWE website hosted by Coventry University (URL: <http://www.coventry.ac.uk/research/research-directory/art-design/british-academic-written-english-corpus-bawe/research-/>) offers a list of links to other studies that have independently made use of the whole BAWE corpus or sub-corpora from it. Only a very small number of these use sub-corpora from BAWE to investigate disciplinary difference. A more frequent use of BAWE data seems to be in investigation of the writing of particular non-native speaker cohorts either contrasting sub-corpora within BAWE, comparing a sub-corpora of BAWE with a learner corpus or utilising BAWE as a reference corpus (e.g. Allen, 2009; Chan and Tan, 2010; Chen and Baker, 2010; Lee and Chen, 2009; Ebeling, 2011). Another use of the BAWE corpus data has been in studies comparing student and professional writing within particular disciplinary fields (e.g. Henderson and Barr, 2010; Breeze,

2011). Four studies listed that do involve contrastive disciplinary analysis are those of Thompson (2009), Bruce (2010), Durrant and Mathews-Aydinli (2011), and Li and Wharton (2012).

Thompson (2009) examines the texts of History and Engineering students in the three years of their undergraduate study with respect to how both discipline and individual writer traits shape these texts, focusing on use of first person *I*, the most frequent *n*-grams, and 'it+linking verb+ADJ+that' and 'it+linking verb+ADJ+to' patterns. One interesting conclusion he draws is that the wider range of genres an Engineering student is required to write means they do not have the same opportunity as a History student to develop disciplinary identity through repeated practice of one genre (p. 80):

From the point of view of identity, it could be argued that the History student builds a fuller identity as a historian through the many experiences of writing essays, while there is less opportunity for Engineering students to develop the range of identities that are required of reflective, professional and laboratory texts.

Bruce undertakes a genre analysis of two small samples of undergraduate writing from Sociology and English Literature respectively (2010) taken from the BAWE corpus. He looks at what he equates to moves in the introductions finding that introductions in both disciplines usually contained the same two, *context* and *outline*, (pp. 157-160) but that in the samples from English literature this was done 'in a considerably more abbreviated way' (p. 160). In terms of his category of 'cognitive genre', as outlined above in section 2.1.4, a greater spread in terms of employment of the four cognitive genres occurred in the sociology samples, while in the English literature samples the cognitive genres of 'explanation' and 'recount' were heavily used whilst there was just one instance of 'discussion' and none of 'report' (pp. 160-162). Bruce also examines metadiscourse use and finds that the sociology samples made more use overall of metadiscoursal mapping both at the beginning of texts and throughout, while essays in English Literature 'appeared to assume greater reader responsibility' using fewer instances of metadiscourse for

explanation of an essay's shape as well as less explicit reader direction towards argumentation points (p. 162). However, in both disciplines he finds that the most frequently used metadiscourse device is 'evidentials', either paraphrases or direct quotes with citations (p. 161), and argues that although there were more instances of the metadiscoursal device of 'frame markers' signalling 'rhetorical shifts' in the Sociology essay texts, these were nonetheless relatively infrequent (p. 163). Durrant and Mathews-Aydlini (2011) analyse the introductions of master's-level essays from eight Social Sciences disciplines as compared to introductions in research articles from equivalent disciplines, looking particularly at the use or non-use of the 'indicating structure' (IS) step. They find that while there is evidence in the professional research writing of strong disciplinary preference for inclusion or not of this step, such specialisation is not exhibited in student writing (pp. 70-71).

Li and Wharton's study, which uses BAWE to create two of four sub-corpora, usefully includes an institutional dimension. In a contrastive analysis of metadiscourse use by native Mandarin speakers undertaking English-medium study at undergraduate level across two disciplines, Literary Criticism and Translation Studies, and across two university institutions, one in the UK and one in China, Li and Wharton (2012, p. 345) found that whilst both disciplinary and institutional influences were evident, the effect of contextual factors may be stronger than that of disciplinary factors. In terms of contextual differences, for example, they found that there was a significant difference in use of 'transition markers' with noticeably more of these used in the UK context, and also differences in choice of marker, with the simpler *but* being favoured in the Chinese context in contrast to greater use of *however* and *therefore* in the UK context (p. 351). In terms of within-context disciplinary variation in the case of the UK context there were significant cross-disciplinary similarities in terms of levels of interactive metadiscoursal resource use, but differences in levels of interactional resource use with proportionately more hedges and attitude markers used in Literary Criticism than in Translation Studies (p. 353).

Recently, Durrant has undertaken two studies of texts across the BAWE disciplines one focusing on high frequency vocabulary items (2014) and the other on recurrent four-word sequences (2015). In the former study he creates high-frequency vocabulary lists delineated by both discipline and level analysing the degree to which they overlap, and then conducts a cluster analysis to determine potentially broader groups which have shared vocabulary needs. He finds 'substantial variation between disciplines' with most disciplines being 'relatively internally homogenous, and so good candidates for teaching units' (p. 328). In the latter study Durrant (2015) lets groupings emerge from analysis of variation in n-gram use across all individual writers within the BAWE corpus. This analysis reveals a primary distinction between 'hard' subjects, science and technology, and 'soft' subjects, humanities and social sciences as well as two other groupings, life sciences and commerce. Interestingly, with regards to the focus of my own project, Durrant finds within the soft Humanities and Social Sciences grouping, the disciplines of Law, History and Politics cluster together to form a strongly associated sub-grouping (p. 11). Durrant conducts a further qualitative analysis of the lexical bundles - n-grams with high frequency across many authors' texts - within the 'hard' and 'soft' groupings in an attempt to identify their distinctive characteristics. Although this is not the place to list all of these characteristics, Durrant finds, for example, that writing in the soft subject grouping is characterised by 'a focus on abstract concepts' an emphasis on 'unique autonomous agents in processes that are difficult to control' while writing in the hard subject grouping is characterised by 'a focus on the physical world' and an emphasis on 'the role of passive, interchangeable, instruments in processes that are tightly controlled by the researcher' (p. 26).

Beyond studies that draw on BAWE data, the work of Cortes (2004), North (2005) and Gimenez (2012) are further examples of investigations of disciplinary difference in student writing. Cortez (2004) identifies the most frequent lexical bundles in corpora of research articles in history and biology before investigating use of these features in student writing from the same disciplines. She finds the bundles are 'rarely used' by students in their writing and that on occasions that certain bundles were used by students, this use was different to that of professional

academic authors (p. 397). Taking an SFL approach North (2005) contrasts the use of Theme in writing produced for a single Open University course in the history of science by students enrolled in arts programmes and students enrolled in science programmes. She finds the arts students generally perform more successfully on this course possibly because of their use of 'themes which framed the discussion as a matter of interpretation rather than fact' and argues that her results 'support the hypothesis that students' writing is shaped by their disciplinary background' (p. 431).

Gimenez (2012) examines undergraduate writing in nursing and midwifery, and finds that the two disciplines treat knowledge, criticality, evidence and impersonality in noticeably different ways arguing that these features are context specific and defined epistemologically in the same way as academic writing is within higher education. His study conducted over two years at a London institution is largely ethnographic in approach involving a questionnaire, semi-structured interviews with both lecturers and students, and two discipline-specific focus groups. There is a small textual component involving analysis of a few sample texts the participants themselves selected from their own writing which they considered to represent their programme's demands with regard to academic writing (p. 406). These texts are not subjected to linguistic analysis but linked to interview data in terms of what the content of the texts reveal. He finds knowledge in nursing to be framed 'within the positivist model based on scientific evidence' (p. 412), with 'criticality ... mainly framed as 'problem-solving' (p. 413), whereas midwifery follows a 'constructivist approach to knowledge creation' in which 'criticality is constructed as a multifaceted attribute which does not only involve problem solving but also critical thinking, critiques of the theories and practices of midwifery, and an understanding of ideology and social change'(ibid).

To summarise, in comparison to research investigating professional disciplinary writing, there is still less research into the disciplinary characteristics of student academic texts. Studies drawing on the BAWE corpus data, particularly Nesi and Gardner's (2012) survey of genres across the disciplines and Durrant's (2015) use of

n-grams and lexical bundles to delineate meta disciplinary groupings and their distinctive characteristics, represent major progress with regards to our understanding of the nature of student academic texts and how this is influenced by discipline. However, there is considerable room for further research into student writing and discipline. There are still very few studies which have contrastively investigated lexico-grammatical level features of writing within specific disciplines/degree programmes. Apart from the work of Hardy and Römer (2013) and Durrant (2014, 2015), most of the studies described above have drawn on pre-existing linguistic frameworks or categories, developed in the study of professional academic writing, with which to approach analysis of target student texts. Therefore, it can be confidently argued that there is both room and a need for further research into the potential disciplinary or other influences on the features of student texts, and for research that seeks to uncover potentially 'new' features of student writing by deploying methods/approaches that do not rely on a priori frameworks.

2.3. Corpus linguistics and disciplinary discourse analysis

The majority of studies discussed above investigating disciplinary characteristics of both professional and student writing have drawn in different ways upon approaches in corpus linguistics. Corpus linguistics approaches have obvious advantages in terms of the speed at which large amounts of linguistic data can be processed, the ease with which this data can be manipulated by researchers, and the greater reliability afforded by computer as opposed to manual human processes (McEnery and Wilson, 2001, pp. 16-17, McEnery, Xiao and Tono, 2006, p. 6). Corpora are designed to be 'principled' collections of 'naturally occurring' language (Conrad, 2002, p. 76); compilers of both 'larger corpora' used to study language use at a broad, relatively generalised level and 'specialised corpora' used to study specific genres within specific contexts (Connor and Upton, 2004, p. 2) attempt to represent as accurately as possible the target language type with regards to the samples or whole texts included in their corpora.

A broad consensus appears to exist among contemporary discourse researchers with regards to the centrality of corpora of authentic texts to genre and discourse analysis. However, there is a plurality of perspectives when it comes to why and how corpora and corpus tools are employed. In this section I will firstly examine four overlapping areas of debate within corpus linguistics which have implications for the methodological choices made by discourse researchers. These debates regard the nature of corpus linguistics, the relative advantages of a 'corpus-based' or 'corpus-driven' (Tognini-Bonelli, 2001) approach, the extent to which corpus linguistics provides an 'empirical' approach, and the extent to which/the particular way corpus tools are useful for discourse analysis. I will then examine one particular corpus tool employed in my project, keywords, and the pertinent issues and debates regarding this tool.

2.3.1 The nature of corpus linguistics

Corpus linguistics is by no means a homogenous field of scholarship. McEnery and Hardie (2012, p 1) state that '[w]hile some generalisations can be made that characterise much of what is called 'corpus linguistics', it is very important to realise that corpus linguistics is a heterogeneous field'. This heterogeneity is generally viewed positively with, for example, Teubert (2005, p. 13) arguing that '[o]nly if the discourse of corpus linguistics remains controversial and pluralist will there be any progress'.

There is contention at the level of the term 'corpus linguistics' itself. A broad divide exists in terms of how corpus linguistics is conceptualised between those, on the one hand, who see corpus linguistics solely or primarily as a 'tool', 'method' or 'methodological approach' (e.g. McEnery and Wilson, 1996; Meyer, 2002; Bowker and Pearson, 2002; McEnery, Xiao and Tono, 2005; McEnery and Gabrielatos, 2006), and those, on the other hand, who see corpus linguistics as having 'theoretical' or 'philosophical'/'paradigmatic' status (e.g. Leech, 1992; Stubbs, 1993;

Tognin-Bonelli, 2001; Mahlberg, 2005; Teubert, 2005). McEnery and Hardie (2012, p. 147) maintain that the latter perspective is held exclusively by scholars working within what they term the 'neo-Firthian tradition' (which will be explained in the next section).

Articles published in the 2010 Special Issue of the *International Journal of Corpus Linguistics*, 'The Bootcamp Discourse and Beyond', are useful for an understanding of the key points of contention with respect to use of corpora between those who see themselves as true/pure corpus linguists, represented in the issue by Teubert (2010a, 2010b) and Louw (2010), and those coming to corpus linguistics from different linguistic traditions, particularly cognitive linguistics, such as Gries (2010). The origin of this issue was an online debate on Corpora List in 2008 regarding the nature of corpus linguistics as either primarily 'methodological' or primarily 'disciplinary' and how or, indeed, *whether* researchers coming from areas other than corpus linguistics should be employing corpora in their work. The fundamental difference from which other key differences of perspective expressed in this issue of the journal stem is the difference between the conception of language as substantially an internal psychological/cognitive process (e.g. Gries, 2010) and the conception of language as a completely external socially constructed phenomenon (e.g. Louw, 2010, Teubert, 2010a, 2010b).

For those who take the latter perspective, analysis of corpora provides the sole source and generator of linguistic theory and therefore 'corpus linguistics' is a linguistic discipline. Teubert argues that all knowledge of language comes from the sum of language that has been produced in context: '[l]anguage for corpus linguists is what takes place between people, it is discourse taking place between members of a discourse community' (2010a, p. 356). Both Teubert and Louw see the use of 'corpus linguistics' by researchers from cognitive traditions as a subversion of the term and misappropriation of corpus data to artificially sustain intuitive models of language: Louw criticises the 'mentalisation of corpus [by cognitive linguists] rather than reduction by corpora of dependence on intuition, introspection, connotation and the plethora of other cognitive terms' (2010, p. 346); Teubert argues 'the label

corpus linguistics has ... been hijacked by theoretical linguists of all feathers' and that '[w]hat used to be and still is for some of us a radically different, a new way to look at language, namely as discourse, not as a quasi-mechanical system, has been foreshortened to a bunch of methods, a toolbox ...' (2010a, pp. 356).

In contrast, the cognitive linguist, Gries, contends that there are in fact many commonalities to be found between social and cognitive linguistics/psycholinguistic approaches to language (2010, pp. 334-336), and argues for a closing of the gap and 'fruitful relation' between them (ibid., p. 338). He argues that Teubert's corpus linguistics (CL) is 'extremely compatible with developments in CogLing/Construction Grammar and with some psycholinguistic theories/models, and that these theories can help CL answer *why*-questions in a better way' (ibid.). He therefore proposes that corpus linguistics should 'assume as the main theoretical framework within which to explain and embed our analyses a psychologically informed, (cognitively-inspired) usage-based linguistics' (ibid.). Wilks is in accordance with Gries in arguing that Teubert and those allied with him, 'corplingers' (Wilks, 2010, p. 409), misunderstand/misrepresent other approaches (ibid.) and take too extreme a stance against non-corpus derived theory (ibid.). Other contributors to 'the Bootcamp discourse ...' issue also question Teubert's and Louw's stark dichotomy between true corpus linguistics as discipline and those they see as misappropriating and reducing corpora to 'tools': Mukherjee labels this 'dogmatic' arguing that (2010, p. 372) '[t]he fact that the use of corpora has become part of mainstream linguistics over the past few decades is a tremendous success story' (ibid.) and that corpus linguistics has a 'two-fold' future as both methodology and discipline (p. 376). McEnery and Hardie (2010, p. 390) argue that 'corpus linguistics like linguistics in general is only diminished when it narrows to doctrinaire positions' and that 'a position of open-minded enquiry, adopting methodological and theoretical insights from either tradition as appropriate, is one shared by a great majority of self-identified corpus linguists' (ibid.).

2.3.2. The ‘corpus-based’ versus ‘corpus-driven’ debate

The debate as to whether corpus linguistics is a methodology or discipline or both links to the question of whether a ‘corpus-based’ or ‘corpus-driven’ approach is preferable. This section explains the difference between these two approaches, and outlines, and to extent refutes, an argument questioning whether, as with the methodology-theory dichotomy above, the distinction is as clear as some scholars maintain.

Corpus approaches used in studies referred to in the ‘Discipline’ section above could be argued to fall roughly into two categories, studies in which corpus methods have been utilised in a ‘top-down’ way to investigate pre-determined linguistic features within the target texts, and those which utilise corpus methods to take a more ‘bottom-up’ approach to exploration of language of target texts to ‘see what’s there’. Within the studies of professional disciplinary writing, Hyland’s work (e.g. 2009) is both an example of the former approach in his selection of citation, reporting verbs, and hedging devices as features to contrastively analyse, and of the latter approach in his identification and contrastive analysis of salient lexical bundles across the disciplinary groupings in his corpus of research articles. Groom (2007) and Malavasi and Mazzi (2010), taking keywords as their starting points, are examples of studies which fall wholly into the ‘bottom-up’ category. Within the studies of undergraduate disciplinary writing the majority can be argued to take the former approach applying predetermined categories or frameworks. Nesi and Gardner (2012) take a top-down approach in their employment of Biber’s dimensional analysis as does Gardner (2008a, 2012) in her contrast of disciplinary groupings in terms of the SFL framework. Li and Wharton (2012) investigate the predetermined category of metadiscourse. Although Cortes (2004) takes a bottom-up approach in generating lexical bundles in her professional history and biology corpora, these bundles then become predetermined categories for investigation in the equivalent student corpora. In contrast, Durrant’s (2015) exploration of lexical bundles in the BAWE corpus takes an entirely bottom-up approach.

What have thus-far been labelled as ‘top-down’ or ‘bottom-up’ corpus approaches are defined by Tognini-Bonelli (2001) respectively as ‘corpus-based’ and ‘corpus-driven’ linguistics. She states that ‘the term *corpus-based* is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora become available to inform language study’ (2001, p. 65). In contrast, in *corpus-driven* linguistics ‘the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence’ (p. 84). In this approach theory does not pre-exist, but is generated from the corpus: ‘[t]he theory has no independent existence from the evidence and the general methodological path is clear: observation leads to hypothesis leads to generalisation leads to unification in theoretical statement’ (p. 85). This argument falls strongly on the ‘theory’ side of the debate as to the nature of corpus linguistics, represented more recently by the arguments of Louw (2010) and Teubert (2010a, 2010b) in the ‘The Bootcamp Discourse’ debate discussed in the previous section; Tognini-Bonelli (ibid.) argues that corpus-driven linguistics (CDL) should have the status of a discipline.

The corpus-based versus corpus-driven dichotomy maps to the distinction Hunston (2002, pp. 92-93) makes between using an annotated corpus and using a plain-text corpus (again, this is also in line with Teubert’s (2010, p. 355) argument that ‘annotation presupposes categories not validated by corpus evidence’). She suggests that, because it is not skewed or limited by a priori categorisations which characterise annotated corpora, a plain-text approach has significant relative advantages for an inductive investigation of the phraseology of a word. Conceding that annotation, particularly of an ‘ad hoc’ nature, has uses in conjunction with plain-text work to move ‘between categories and words’, she nonetheless maintains that ‘[a]nnotation should serve the needs of the corpus user, not determine the direction the investigation must take’ (p. 93).

Tognini-Bonelli, Hunston and Teubert can be seen as working within a linguistic tradition, labelled ‘neo-Firthian linguistics’ by McEnery and Hardie (2012), which is

strongly influenced by the work of John Sinclair and associated with the University of Birmingham. Informed by Firth's (1968) pre-corpus-era conception of 'collocation', work in this tradition privileges lexis over traditional grammar in accounts of language with 'the meanings and phraseologies of words in context [being] major concerns of neo-Firthian corpus linguistics' (McEnery and Hardie, 2012, p. 142). Sinclair (1991) brought Firth's idea of 'collocation' into the corpus era by defining it as referring to co-occurrence patterns as observed in corpus data. Louw (1993) and Stubbs (2001) build on the idea of 'collocation' with the respective concepts of 'semantic prosody', the association of a lexical item with either positive or negative meaning in the context of use, and 'semantic preference', the association of a lexical item with words belonging to a particular semantic category co-occurring in a text. Sinclair's (1991) 'Idiom Principle' provides a linguistic framework which enlarges the role of lexis while diminishing the role of grammar or, in other words, diminishing the importance of the 'open-choice principle' which in traditional grammars was seen as accounting for all linguistic behaviour. Hunston and Francis's (1999) 'Pattern Grammar' framework and Hoey's (2005) 'Lexical Priming' framework go even further in that both in different ways '[unify] the description of lexis and grammar largely or entirely' (McEnery and Hardie, 2012, p143). All of the related approaches discussed above deem as legitimate only conclusions about language that emerge from the corpus data and this clearly links to the preference for the 'corpus-driven' rather than 'corpus-based' approach delineated in the previous paragraph.

McEnery, Xiao and Tono (2006) and McEnery and Hardie (2012), however, take issue with the view that corpus linguistics divides between 'corpus-driven' and 'corpus-based' approaches. McEnery et. al. (2006, p. 11) argue that 'the sharp distinction between corpus-based vs. corpus-driven approaches to language studies is in reality fuzzy'. Pointing out that 'corpus-driven linguists do concede that pre-corpus theories are insights accumulated over centuries which should not be discarded readily and that intuitions are essential in analysing data', they argue that consequent use of traditional grammatical categories such as nouns and verbs combined with the fact that 'linguistic intuitions typically come as a result of

accumulated education in preconceived theory' means that 'there is no real difference between the corpus-driven demand to re-examine pre-corpus theories in the new framework and the corpus-based linguists' practice of testing and revising such theories' (ibid., p. 10). McEnery and Hardie (2012, p. 149) conclude that the alternative labelling of 'corpus-as-theory' versus 'corpus-as-method' is more useful in terms of distinguishing the approach of those working within or without neo-Firthian corpus linguistics:

[W]e would argue that the different schools of corpus linguistics are not reliably distinguished – or, alternatively, that their nature is not optimally communicated – by the corpus-based versus corpus-driven distinction, or by the 'top-down' versus 'bottom-up' distinction with which it is often equated. Rather, it is in the contrasting stances on the conceptual status of the corpus and of corpus linguistics – as having theoretical status versus as a linguistic methodology – that truly separates the two schools.

However, whilst the corpus-driven vs. corpus-based distinction may not be as clear cut as is sometimes claimed, McEnery and Hardie's assertion that there is no real difference goes much too far as a counter-argument. Whilst it is obviously impossible to free oneself of all pre-existing knowledge of language theory, there is a very real difference between taking a corpus approach designed to as far as possible let patterns and features emerge, and taking a corpus approach which involves locating pre-determined linguistic features. For this reason, I would argue the distinction between corpus-driven and corpus-based approaches remains an extremely useful one.

Related to the debate above is the question of whether when adopting a corpus approach the researcher regards the corpus methodology/ies employed as entailing a positivist or interpretivist framework for the analysis. The extent to which corpus linguistics provides an empirical, 'hard science' perspective on language and discourse is questionable. Whilst McCarthy (2001, p. 125, cited in Taylor, 2008, p. 181) states that corpus linguistics represents 'cutting edge change in terms of scientific techniques and methods', there is acknowledgement by many of the

limits to which it can be described as a 'scientific' approach. Teubert (2009, paragraph 35) argues that corpus linguistics:

can only partly be seen as a hard science. Its methodology and its tools may well deliver dependable and reproducible results. But these results do not tell us much. They need to be interpreted. Yet this takes us into the human interpretive sciences.

Gries, in contrast, from a 'psycho- and cognitive-linguistically informed' perspective, sees corpus linguistics as 'located, firmly and deliberately, in the social/behavioural sciences' (2010, p. 333).

Baker argues (2006), that whilst more transparency and 'objectivity' is arguably achieved by using a corpus approach, it is not empirically or scientifically objective in nature. Taking into account to Burr's (1995, cited in Baker, 2006) social constructivist argument that objectivity is impossible due to the fact 'the "objective" stance is still a stance', Baker (2006, pp. 10-12) suggests true objectivity cannot in fact be achieved using a corpus approach due to human limitations in terms of the processing information and interpretation of evidence, and also due to conscious and subconscious interpreter bias. Baker, however, does not take an extreme social constructivist position on corpus approaches, but rather argues that '[b]y using a corpus ... we are at least able to place a number of restrictions on our cognitive biases' (p. 12). Thus, corpus research could be argued to be interpretivist but moderately so, with the various statistical tools available to corpus researchers reducing but not completely eliminating interpreter bias.

2.3.3. The usefulness of corpus linguistics to discourse analysis, particularly in ESP

It could be argued that a problem with employing a corpus approach to analyse discourse lies in the fact that, by their nature corpus methods, tend to divorce language from context. Both Hunston (2002) and Baker (2006) discuss this issue with regards specifically to critical discourse analysis. Hunston (2002, p. 110) states that applying corpus techniques to a corpus of whole texts will 'obscure the

character of each text as a text' thereby obscuring context of society and individual author. Baker (2006, p. 7) also sees as potentially problematic the 'broadness' of perspectives derived from corpus analyses of a large volume of texts wherein close reading and understanding of individual texts is sacrificed. These issues can also be seen to relate to researchers working not within critical discourse analysis but ESP. Flowerdew (2005) outlines questions raised about the usefulness of corpus approaches in ESP/EAP with reference to analysis at the level of genre: firstly, its 'atomised, bottom-up' (p. 324) characteristics are seen as at odds with the genre analysis approach which usually starts at the macro-structural level and works downwards (Swales, 2002, cited in Flowerdew, p. 324); secondly, it is argued corpus approaches divorce language from communicative context (Flowerdew, 2005, p. 324) and for that reason the authenticity of language is undermined when it is presented within a corpus (Widdowson, 1998, 2002, cited in Flowerdew, p. 324). Flowerdew sees 'tagging systems for coding the generic 'move structures' of ESP texts under investigation' (p. 325) as a solution to the former issue. With regards to the latter issue of de-contextualisation, she advocates 'working with small, specialized corpora ... where the analyst is probably also the compiler and does have familiarity with the wider socio-cultural context in which the text was created' (p. 239). She argues that the 'compiler-cum-analyst' can then assume the role of 'a kind of mediating ethnographic specialist informant to shed light on the corpus data'.

There is also an argument that the lexico-grammatical level at which corpus linguistics tends to operate is potentially useful in shedding light on discourse features of target language/texts. Baker (2006, pp. 13-14) argues, again from a critical discourse analysis perspective, that discourse has an 'incremental effect' and thus that a significant amount of textual data needs to be examined to identify patterns in ideas and meanings which only emerge in this examination of a large number of texts. He quotes Stubbs (2001, p. 215, cited in Baker, 2006, p. 13): 'repeated patterns show that evaluative meanings are not merely personal and idiosyncratic, but widely shared in a discourse community'. Stubbs' and Baker's arguments could be seen as equally relevant to ESP studies of the

professional/academic communication of particular discourse communities; sentence-level linguistic features that are identified to be frequently repeated across instances of a particular genre shared by a particular discourse community are likely to be revealing of salient characteristics/concerns of that particular discourse. Indeed, Hunston (2002), in reviewing different approaches to analysis of lexico-grammatical level of particular disciplinary discourses, comes to the following conclusion (p. 204):

The techniques of corpus investigation are ideally suited to examining specific, fairly homogenous discourses. Specialised corpora can be compiled relatively easily and connections can be made between the phraseology of a discourse and the ideology of the discourse community. The results are clearly applicable to the needs of those seeking to be socialised into that community.

2.3.4. The usefulness of keywords to discourse analysis

The 'keywords' method has proved a popular tool for researchers working with specialised corpora (Hunston, 2002, p. 68) providing as it does an objective starting point for further corpus analysis (Groom, 2007, p. 51). This method has been made popular by Mike Scott's keyword programme, part of the corpus analysis software suite Wordsmith Tools (e.g. Scott, 2001), and involves identifying statistically significant lexical items within the corpus which is the object of research focus, *the research corpus*, in comparison to a *reference corpus* which is larger and more general. In terms of analysis of specialised discourse, the method could be seen as particularly useful due to its argued ability to objectively access important themes within a corpus; Scott and Tribble (2006, pp. 55-56) maintain:

keyness is a quality words may have in a given text or set of texts, suggesting that they are important, they reflect what the text is really about, avoiding trivia and insignificant detail. What the text "boils down to" is its keyness, once we have steamed off the verbiage, the adornment, the blah, blah, blah.

Baker (2004) reviews the strengths and weaknesses of the keywords tool with respect to critical discourse analysis. Whilst acknowledging the usefulness of its

speed and objectivity with regards to identifying lexical items that are unusually frequent or infrequent, he argues that 'it is essential to realize that a keyword list only provides the researcher with language patterns which must be interpreted in order to answer specific research questions' (2004, p348). He highlights the potential danger, when two target corpora are compared with each other to generate keyword lists, of an overemphasis on difference and neglect of similarities and ways in which the same word has different patterns of usage between corpora. He also points out that on a keyword list the 'strongest words tend to reveal most obvious differences' that 'we could probably have made an educated guess at in advance' (2004, p350). Overall, his argument is that an initial keyword list should not be taken at face value and that there are a range of further steps that can be taken, such as semantically grouping infrequent keywords, analysing dispersion across files to identify 'key keywords', words that are key in each individual file that constitutes a corpus, or undertaking concordance and collocational analyses, in order 'to obtain a more accurate picture of how keywords function in texts' (2004, p357).

Groom (2010, p. 60) argues for the advantages of a keyword approach to the discourse analyst for finding 'genuinely new' features of the target discourse; he states that the simplicity of the algorithm at the heart of the keyword procedure 'means that it is entirely unencumbered by previous linguistic theories or descriptions' and that the procedure 'has a well-attested knack of unearthing features and trends in corpus data that would be difficult or impossible to observe by more conventional methods' (ibid.). Groom (2010, pp. 71-72) also argues for the advantage of keywords over other bottom up corpus tools such as lexical bundle or word cluster analyses with regards to revealing repeated patterns of meaning within a corpus. He argues that, for instance, a random concordance sample for *of* will not only reveal the fixed expressions generated by lexical bundle or cluster analyses but also 'phraseologies which, while just as conventionalised ... are much less rigidly formulaic, and may therefore be under-emphasised or overlooked altogether by automatic routines' (p. 71). This argument links to that in the next

section regarding the possible advantage of focusing on grammatical rather than lexical items on a keyword list.

2.3.5. Choice of items for further analysis from a keyword list: an argument for focus on 'grammatical' items

A corpus designed to represent the linguistic features of a specialised discourse is likely to generate at the very least many hundreds of words which obviously cannot all receive close analytical attention; a challenge facing discourse researchers who take the keywords procedure as a starting point is, therefore, how to determine in a principled manner what items from a keyword list should be selected for the focus of further analysis (Groom, 2010, p. 60). Researchers may choose a 'top-slicing' approach, in which only the first few words on a keyword list are focused on, or may choose to focus on particular categories of words, discarding the rest (ibid., pp. 60-61). Groom argues that the former approach is 'an ad hoc solution at best' describing the 'frustration' of researchers 'noticing keywords lurking just below their chosen cut-off point that look much more interesting than many of the words above it' (ibid., p. 60). Regarding the latter option, he cites Baker's (2006) observation that 'the default option ... is to discard from the outset all of the so-called "grammatical" words' (ibid., p. 61). Grammatical words have generally been seen to be less interesting to analyse than lexical words (2006, p. 127). They have also been perceived as being less important to the 'aboutness' of the corpus; Scott and Tribble (2006, p. 60) explain the keyness of 'grammatical' items as being related to 'style' rather than 'aboutness'.

Groom (2010, p. 61), however, argues for the opposite perspective, making a case for 'discarding all of the open-class items in a keywords list as a preliminary step, and instead focusing on the closed class keywords that remain'. Extremely few scholars have previously taken this approach; Gledhill (2000) focused on closed-class words in a genre-oriented study of the differences between the different sections of medical research articles (at the time of writing the author of this project is not aware of any studies of specialised discourse by scholars other than

Gledhill or Groom which have taken closed-class items from a keyword list as a starting point). Groom argues for the usefulness of closed-class keywords on a number of grounds. Firstly, a focus on closed-class words means 'a compact and tractable list of items for analysis from the outset' eliminating the need for arbitrary top-slicing (2010, p. 70). Secondly, Groom suggests such a list may be 'a more fruitful as well as more manageable set of starting points for concordance analysis' due to the fact it entails focus on 'the commonest words in virtually all corpora' and therefore coverage of a far greater proportion of the whole corpus than would even a larger set of open-class items (ibid., p. 71); he argues (ibid.):

If it is true that "the majority of text is made of the occurrence of common words in common patterns", as Sinclair (1991: 108) suggests, then it is arguably preferable to select the commonest of these common words as the empirical basis of a corpus-driven discourse analysis.

Finally, as already mentioned in the section above, concordance analysis of patterns surrounding closed-class words allows a researcher to identify repeated meanings in the phraseology of these items, 'semantic sequences' (Hunston, 2008) (explained fully in the next paragraph), which would not be picked up by the more limited lexical bundle or word cluster tools because of their 'very heterogeneous surface realisation' (Groom, 2010, pp. 72-73). Groom argues that such patterns of repeated occurrence are much more likely to be found with focus on closed-class rather than open-class words (ibid., pp. 72-73):

the analyst would have to study, sift through and cross classify the phraseological profiles of a forbiddingly large number of individual open-class items in order to generalise a single semantic sequence ... [whereas] [c]oncordancing closed-class words ... allows the analyst to proceed directly to identification of the underlying (and frequently expressed) commonalities of meaning among superficially very different looking sequences of linguistic elements.

Phraseologies in a corpus which can be grouped in terms of their meaning but which do not necessarily share an homogenous form are termed 'semantic sequences' by Hunston (2008) who provides the following definition (p. 171):

Semantic sequences are recurring sequences of words or phrases that may be very diverse in form and which are therefore usually characterised as sequences of meaning elements rather than formal sequences.

She argues semantic sequences are more useful in analysis of specialised corpora than of general corpora (ibid., p. 272) and that grammatical words 'are the best starting point for identifying semantic sequences in specialised corpora' as 'such sequences identify "what is often said" in those corpora' (ibid.). Citing the research of Gledhill and Groom, she argues for the usefulness of grammatical words to investigations of disciplinary features of texts: '[w]hereas lexical words are a good place to find subject matter of the discipline, grammar words function to find its epistemology' (p. 293). She contends (p. 292):

"small words" ... reveal a surprising amount about the epistemology and ideology of disciplines, because they reveal phraseologies that are linked to recurrent meanings and functions rather than subject matter.

To summarise, this literature review has outlined two broad perspectives on corpus linguistics. The first sees corpus linguistics as a discipline informed by a social understanding of language in which all linguistic theory is generated through analysis of real recorded instances of social use of language. This first approach favours a bottom-up, or corpus-driven, approach to corpus data which generally involves working with unannotated corpora. Advocates of this approach/perspective tend to be critical of corpus-based approaches which test pre-existing non-corpus theories about language. The second perspective on corpus linguistics sees it more as a range of methodological options that can be both employed to generate entirely new insights about language as well as to test existing theory, and also sees the two perspectives, corpus as discipline and corpus as methodology, as overlapping rather than dichotomous. For the particular applied purpose of my project, that of gaining a 'new' perspective on student disciplinary writing at the lexico-grammatical level, the 'corpus-driven' starting point involving unannotated corpora is well suited.

This section has also discussed where a corpus linguistic approach might fit paradigmatically on the cline from empirical to interpretivist, and suggested that whilst corpus techniques serve as far as possible to reduce interpreter bias, ultimately the data objectively, reliably generated by the computer tool becomes the subject of human interpretation. With my project I can therefore be argued to be working within a moderate interpretivist paradigm in which interpreter bias is usefully reduced, but by no means eliminated, through the use of corpus tools.

Arguments have been outlined regarding the usefulness of lexico-grammatical level corpus analysis to understanding of the discourses produced by particular discourse communities. It is argued that by revealing repeated patterns of meaning at sentence level across large numbers of texts, corpus approaches can reveal the shared preoccupations of a discourse community. The usefulness of the keyword tool as a starting point for discourse analysis has been argued on the grounds that keywords are revealing of the 'aboutness' of discourse as represented by a set of texts. The argument that focus on grammatical items from a keywords list is particularly useful in the analysis of disciplinary discourses because they reveal 'recurrent meanings and functions' (Hunston, 2008, p. 292) has also been outlined. In line with this argument, my project takes grammatical keywords as the starting point for investigating potential disciplinary features of the discourse of student undergraduate writing.

Chapter 3: Methodology

3.1. Approach

As set out above in the concluding paragraphs of the literature review, this project takes a bottom-up or 'corpus-driven' approach to analysing the discourse of undergraduate student writing in two disciplinary fields, History, and Politics and International Relations. In this approach the corpora remain unannotated with no a priori linguistic categories employed; the idea is that categories or patterns will emerge from close analysis of the target corpora. Keywords analyses are used as the starting point for further analysis revealing as they do words which are unusually frequent within a corpus and therefore potentially revealing of the 'aboutness' of the texts represented by the corpus. In this project grammatical rather than lexical/content keywords are selected for further close analysis following the arguments of both Groom (2007, 2010) and Hunston (2008) regarding the usefulness of the identification of semantic patterning around grammatical words for developing an understanding of the repeated functions and meanings favoured by a particular disciplinary community. Whilst keywords analyses rest on a statistical calculation, this research is seen to sit within an interpretivist rather than empirical paradigm with corpus tools reducing but not eliminating interpreter bias.

3.2. Research questions

The primary purpose of the study is to investigate within the UK higher educational context notions of 'discipline' or 'disciplinarity' in student, as distinct from professional, academic writing. In order to explore the extent to which any differences identified between the writing students undertake within different UK degree programmes can be attributed to the influence of 'discipline', the study will take into account in its design a potential influence other than discipline, that of particular location/institution. The focus of textual analysis will be at lexico-grammatical level, and the research questions are listed below.

1a. In a comparison between successful third-year undergraduate texts written within the History degree programme and those written within Politics/International Relations degree programme at Institution *X*, what *similarities* can be identified in the lexico-grammatical and semantic patterns occurring around positively key closed-class grammatical items?

1b. In a comparison between successful third-year undergraduate texts written within the History degree programme and those written within Politics/International Relations degree programme at Institution *X*, what *differences* can be identified in the lexico-grammatical and semantic patterns occurring around positively key closed-class grammatical items?

2. To what extent are similarities and differences identified between writing in programmes at institution *X* replicated in a comparison/contrast of successful third-year writing between the equivalent degree programmes within Institution *Y*?

3. What arguments can be made from the findings of the first two questions in terms of the extent to which either 'discipline' and/or local institutional context seem to have a stronger relationship to any similarities or differences found among the lexico-grammatical patterns/phrasology exhibited by the different groups?

3.3. Data selection

The purpose and research questions above shaped the selection of data in terms of the level, nature and quality of writing targeted, the 'disciplines' or degree programmes selected for comparison, and the need to investigate parallel programmes across two institutions. Coursework writing of relatively highly-achieving undergraduate students in their third year in the departments of History and Politics and International Relations (PIR) at two institutions was selected as focus for the study.

The writing of third-year undergraduate students was chosen as opposed to that of first or second-year students following the rationale that by a student's third year they would have ventured as far as is possible at undergraduate level into the discourse community of their department and programme and that therefore at this stage their writing would be an accurate reflection of the influence of any disciplinary and/or institutional socialisation or apprenticeship. The work of relatively highly-achieving third-year students, those achieving largely 2:1s and firsts for their essays, was chosen because this successful performance can be seen as validation in the eyes of the department of their success in terms of the disciplinary expectations of undergraduate students.

Undergraduate as opposed to postgraduate writing was focused on because many students on taught postgraduate degree programmes have studied within a different disciplinary area or in a different national context at undergraduate level; for this reason, their adaption to the discourse community, discipline and institution is likely to follow a steep trajectory within the one year of their master's course, rendering the writing produced by some students at this level possibly less reliably or consistently shaped by practices of their department. Also, there is a considerably smaller cohort at this level meaning it is a much more challenging task to collect an adequate range and quantity of 'good' texts.

Coursework writing as opposed to exam writing was selected as focus for the reason that coursework writing is by nature more in-depth and carefully-constructed than writing completed under timed conditions, and is thus potentially a better reflection a student's knowledge, ability and practices. From a logistical perspective it was also easier to collect than exam writing. The final third-year dissertation was excluded for reasons in line with Alsop and Nesi's rationale for the exclusion of dissertation work from BAWE (2009, p. 74), that this much longer, more polished, multiply drafted piece of work completed under greater guidance from a supervisor, was less characteristic of student writing as a whole.

Writing from degree programmes within the departments of History and PIR was chosen as focus for two connected reasons. Firstly, as discussed in the literature review, relatively few studies of academic writing, either at professional or at student level, have compared writing that sits reasonably closely at the 'soft' end of the disciplinary spectrum and there are compelling arguments (e.g. Nishina, 2010) for the usefulness of such research. History and PIR sit within the soft epistemological field. At Institution X, the researcher's place of work, they fall under the umbrella of 'Arts and Social Sciences' and students have the option of undertaking a joint honours degree across both departments, either in 'Modern History and Politics' or 'Modern History and International Relations'. History and PIR are thus distinct but connected fields and as such a potentially useful focus for identifying finely delineated disciplinary differences. A further reason supporting their choice is the fact that in both areas of study the majority of coursework writing assignments fall under the category of 'the Essay genre family' (Nesi and Gardner, 2012). Only texts from this genre are included in the study and thus distinctions identified between writing in the two areas can more confidently be attributed to discipline and/or institution with the variable of genre as far as possible discounted.

Finally, the study's overall purpose of investigating the extent to which 'discipline' and/or local institutional context exerts an influence on undergraduate writing necessitates collecting data from more than one institution. Without a cross-institutional analysis, differences identified between writing done by students in the two programmes within Institution X could only confidently be argued to be differences between the discourses of those particular localised departments. However, if the same patterns of difference found between Institution X programmes are also found between equivalent departments at another UK HE institution, the attribution of these differences to 'disciplinary' factors becomes much more compelling.

To create a cross-institutional dimension to the study, data from the British Academic Written English (BAWE) corpus (a resource discussed in more detail

below) was usefully employed. Although the BAWE corpus texts were collected from four institutions, holdings for each discipline generally consist of student writing from a single institution (Alsop and Nesi, 2009, p. 74); an examination of preliminary records, held at CAL, of the compilation of the BAWE corpus confirmed that this is indeed the case for Politics and History. ‘Politics’ essays in BAWE were collected from a department named ‘Politics and International Studies’. As can be seen in Table 1 below, a cross-programme, cross-institutional comparison was designed with Institution X being the institution I work within and Institution Y being the institution from which BAWE texts were collected.

| | History | Politics/IR |
|----------------------|----------------|--------------------|
| Institution X | HIST-X | POL/IR-X |
| Institution Y | HIST-Y | POL/IR-Y |

Table 1: Corpus design

3.4. Data collection

3.4.1. Data from the BAWE corpus

The BAWE corpus comprises a set of texts written for assessment purposes by both undergraduate and postgraduate students on taught programmes within the UK higher education system (Alsop and Nesi, 2009, p. 71). The majority of texts within the BAWE corpus are written by ‘native speaker’ students; however, all texts included were required, irrespective of the first language of the contributor, to be assessed as ‘proficient’ by academics from the relevant discipline (ibid.). The rationale behind the construction of the BAWE corpus was to provide a resource for research into features of successful discipline-specific student writing, providing ‘strong quantitative insights into student writers’ use of grammar, lexis, and discourse patterns across disciplines’ (Nesi et al., 2004, p. 443, in Alsop and Nesi, 2009, p. 72), for the purpose of informing academic writing tuition (pp. 71-72). Prior large-scale studies of writing in a university context such as the PERC Corpus of

Professional English, the TOEFL 2000 Spoken and Written Academic Language Corpus, and the International Corpus of Learner English, had focused, in the case of the first two listed, on 'published or publicly accessible texts' or, in the case of the last, 'learner essays on general academic topics ...designed primarily to monitor non-native-speaker errors and the processes of language acquisition, rather than the development of academic literacy skills and disciplinary knowledge' (ibid.) For this reason the BAWE corpus serves to fill a gap in terms of corpus resources, its aim being 'to enable the identification and description of student writing genres across disciplines and at different stages of academic development' (p. 72). Before the completion of the Michigan Corpus of Upper-level Student Papers (MICUSP, 2009) it was 'the only formally planned and archived corpus of its kind' (p. 72).

Texts for the BAWE corpus were collected between 2005 and 2007 primarily from Warwick University, Reading University, and Oxford Brookes University with some collected from Coventry in the later stage of the collection process (p. 73). A 'matrix' consisting of four disciplinary categories, Arts and Humanities, Life Sciences, Physical Sciences and Social Sciences, and four levels of study, from first to third year undergraduate and fourth-year for one-year post-graduate master's programmes, was used to structure the corpus organising it into 'sixteen cells of approximately equal size' for the sixteen individual disciplines included (ibid.). Disciplinary categories were chosen to enable easy comparison with important corpora of spoken academic English, the Michigan Corpus of Spoken Academic English (MICASE) and the British academic Spoken English (BASE) corpus which use very similar categories (ibid.). Assignments, excluding master's theses, with both a formative and summative purpose which had received a mark of at least sixty per cent from their source department were collected (p. 74). Each text included in the corpus was assigned a 'genre family' label from the categories of case study, critique, design specification, empathy writing, essay, exercise, explanation, literature survey, methodology recount, narrative recount, problem question, proposal, and research report (Nesi, 2008). Information included regarding the first language, number of years of UK secondary education, whether the assignment was categorised as 'merit' (equivalent to upper second class) or 'distinction' (equivalent

to first class) can be used by researchers accessing the BAWE to create a sub-corpus fitting their particular requirements (ibid.).

3.4.2 Data from my own institution

Texts from third-year students in disciplines at my own institution were collected in order to construct two disciplinary sub-corpora equivalent to those from BAWE. As discussed above my aim was to collect texts from students who were performing relatively highly by their final year of study, and I achieved this with the large majority of texts collected having grades of at least 60% with a very large proportion of these having grades in the 'high 60s' and a significant minority in the 70s. This means that in terms of the sub-corpora from my institution being composed of 'good' student work, there will be equivalency with BAWE.

Collecting data from my institution was both challenging and time-consuming. I made the decision to start relatively early with this process based on anticipated difficulties learned from experience as an EAP tutor attempting to communicate with academics in departments needing EAP support and also seeking out example student writing to better inform development of teaching materials. Academics, although usually positive about and supportive of what I was trying to do, very often proved unreliable in terms of following up on promises of cooperation or responding in a timely fashion to emails. I had also found it very difficult to obtain single examples of writing from students, so expected that it would be a considerable challenge to collect student work on the scale that I needed.

The challenges in collecting such data are described with reference to the BAWE project by Alsop and Nesi (2009) who say of the pilot corpus for this project that it 'illustrated the difficulty of collecting a representative selection of work from a shifting student population, who produced varying amounts of writing at various stages of the year, and who had relatively little incentive to cooperate with our research agenda' (pp. 72-73). They describe an evolution in the publicity strategies

they employed over the course of collection of texts for the BAWE (pp. 77-80) and also the compromises that had to be made due to failure to achieve original targets in some areas which included the need to go beyond one institution to collect enough texts for certain disciplines (p. 74), and the need to ease restrictions in terms of how many texts a single student could submit (pp. 75-76).

My data collection process spanned just over a year taking in the finishing third-year cohorts of 2011 and 2012, and, due to very little success in 2011, I changed my strategy in 2012. To find students from PIR I initially emailed a senior academic whom I had communicated with reasonably regularly in relation to an ESAP course I ran for international master's students. The email was forwarded to administrative staff to disseminate to students and I heard nothing more. Alsop and Nesi (2009, p. 77) state that '[a]n e-mail with departmental endorsement was assumed to hold more weight than an e-mail directly from [the BAWE team]'. I would agree but go further in arguing that unless students are communicated with directly by an academic whom they respect, ideally in person rather than via e-mail, they are unlikely to respond (I collected texts from Marketing students in the School of Management, which I will not be using for this project, and this proved considerably easier because of a supportive academic who selected and communicated with students directly). I obtained one student volunteer through my contact in PIR who was directly introduced to me at an orientation event I attended as part of my job.

For History, I initially managed to find two participants quite quickly through a personal contact who was studying as an adult student in the department. I then got in touch with an academic within the department who had been recommended as likely to be supportive, and over the course of three months exchanged emails, met for coffee to explain my project, attended one of his tutorials to meet four of his students, all of whom agreed to take part, followed up on this meeting by emailing those four students on three separate occasions, before having just one of these students ultimately participate in the project.

I did not initially offer any kind of payment for participation in the project. However, having only collected texts from two students in History and one in Politics in almost a year I decided to offer book vouchers as payment for participation. I also changed tack and contacted the student executives of both the History and PIR societies on campus. Both these changes in approach proved fruitful. All students I emailed responded and then, through some of these students, I was able to get recommendations for others whom they thought would be suitable. I sent three waves of emails to all students I had contact details for. The gap between spring and summer term seemed to be the most effective timing in terms of getting responses, these then tailed off through May and early June during the examination period, with a few final students responding in the period directly after exams. Overall, roughly half of those I emailed participated and donated texts.

Collecting data from each student involved setting up a face-to face meeting for an interview and students generally emailed me their essays or downloaded them from Turnitin from my office computer on the day of the interview. Collecting all related information, such as grades and course names was not always easy often involving follow-up emails not all of these responded to. There are two students I do not have specific essay grades for, but am confident that both these are '2:1/1' students; both of these students were recommended to me by academics as 'good students'.

Having collected the texts from students at my own institution in Microsoft word form, I then converted them to plain text files and 'cleaned' them for use in the sub-corpora. Cleaning involved placing chevrons around Harvard references, finding and deleting footnote numbers within the texts, and deleting titles, footnotes and references/bibliographies.

3.5. Creation of research corpora

The corpora created for this study are ‘specialised corpora’, designed and constructed to be representative of a particular type of text the specialised language of which they will be used to investigate (Hunston, 2002, p. 14; McEnery, Xiao and Tono, 2006, p. 15). With texts from my own institution and the BAWE corpus texts discussed above I created four sub-corpora, *HIST-X*, *POL/IR-X*, *HIST-Y* and *POL/IR-Y*. Plain text files of all third-year texts from the Essay genre family (Nesi and Gardner, 2012) were extracted from both the History and Politics holdings of BAWE to create the two *Institution Y* sub-corpora. The HIST-Y sub-corpus consists of 30 third-year essays, contributed by 12 students, with a total word length of 88,228 words. The longest third-year essay is 9037 words and the shortest 1785 words (there are three texts approximately 9000 words long but the majority are between 2000 and 3000 words in length). The POL/IR-Y sub-corpus consists of 13 third-year essays, contributed by six students, with a total of 45,689 words. The longest essay is 5445 words and the shortest 1589 words. Two students contributed to both the History and the Politics sub-corpora; 0244 (in BAWE documentation, each student has a number to which an alphabet letter is added to differentiate different texts contributed by the same student) contributed one second and four third-year essays to Politics and three third-year essays to History, and 0147 contributed one second-year essay to each discipline. Of all third-year assignments within BAWE only one was excluded from the History sub-corpus, a ‘critique’. Two were excluded from the Politics sub-corpus, a ‘critique’ and a ‘literature review’ written by third-year students, for the reasons of genre equivalence discussed above.

The HIST-X sub-corpus consists of 32 third-year essays, contributed by 7 students, and totalling 96,711 words of third-year writing. The longest essay is 5479 words and the shortest 2049 words. A direct comparison with the HIST-Y sub-corpus is shown in the table below. Due to the fact that fewer students contributed to the HIST-X sub-corpora, the influence of individual style will have to be born in mind and cross-checked for. However, a potential strength in the RHUL sub-corpus is the

fact that most students contributed essays evenly from across their programmes enabling a potentially more accurate picture of the range of writing they are required to do. Details of the HIST-X and HIST-Y corpora content are contained in Table 2 below.

| | HIST-X | HIST-Y |
|--|----------------------|----------------------|
| 3 rd year - total words | 96,711 | 88,228 |
| 3 rd year – no. of essays | 32 | 30 |
| 3 rd year maximum and minimum length | Max: 5479; Min: 2049 | Max: 9037; Min: 1785 |
| No. of students contributing 3 rd year essays | 7 | 12 |

Table 2: HIST-X and HIST-Y sub-corpora

The POL/IR-X sub-corpus consists of 33 third-year essays, contributed by 7 students, totalling 92,662 words. The longest essay is 3914 words and the shortest 1647 words. A direct comparison with the POL/IR-Y sub-corpus is shown in the table below. Unlike the situation with the history sub-corpus there is reasonable equivalency between the POL/IR-X sub-corpus and the POL/IR-Y sub-corpus in terms of the number of students contributing. However, as is potentially the case with the RHUL history sub-corpora, the POL/IR-X equivalent possibly benefits in terms of the insights that could be gained from the fact many students contributed assignments evenly from across their programmes. Two students contributed essays to both HIST-X and POL/IR-Y sub-corpora. Details of the POL/IR-X and POL/IR-Y corpora content are contained in Table 3 below.

| | POL/IR-X | POL/IR-Y |
|--|---------------------|----------------------|
| 3 rd year - total words | 92,662 | 45,689 |
| 3 rd year – no. of essays | 33 | 13 |
| 3 rd year maximum and minimum length | Max: 3914 Min: 1647 | Max: 5445; Min: 1589 |
| No. of students contributing 3 rd year essays | 7 | 6 |

Table 3: POL/IR-X and POL/IR-Y sub-corpora

A summary of both cross-discipline/programme and cross-institutional corpus data can be seen in table 4 below.

| | History | Politics/IR |
|----------------------|--------------------------|----------------------------|
| Institution X | X HIST (96,711 words) | X POL/IR (92,662 words) |
| Institution Y | Y HIST (88,228 words) | Y POL/IR (49,689 words) |

Table 4: All four sub-corpora

In an ideal world, all four sub-corpora would be of equivalent size and would also be matched in terms of numbers of texts and numbers of contributors. In the real world, such exactitude is rarely possible – even the BAWE project, with all its human, institutional, and financial resources, was unable to achieve it. The fact that, for example, the POL/IR-Y component is smaller than the other three sub-corpora is a constraint which needs to be accepted. In the following section, I discuss issues of equivalence in more detail.

3.5.1 Design considerations: size, representativeness and balance

With regards to corpus size, Hunston (2002, p. 26) states that ‘arguments about size tend to be academic for most people. Most corpus users simply make use of as much data as is available, without worrying too much about what is not available’. The corpora created for this study are relatively small when compared to those that have drawn on academic texts such as theses and research articles available in the public domain. Groom (2007, p. 71), for example, compiled four corpora ranging between 1,011,238 and 4,057,104 words in size for his study of the research article and review genres in History and Literature, and Nishina (2010, p. 62) compiled corpora of 2,667,000 words and 2,668,679 words respectively for his study of research articles in Applied Linguistics and Business. The relative smallness is largely due to the difficulties involved in collecting student texts (discussed above).

In collecting data from my own institution, I have aimed for an equivalent size in terms of the number of texts collected for each year of each discipline for BAWE. This has been achieved, but, as discussed above, significantly fewer students contributed texts to the History corpus from my own institution. There is better equivalency in terms of number of students contributing for the PIR/Politics corpora; however, this is due to the fact that there was a significant shortfall in the number of Politics texts collected for BAWE which creates a further issue in that the POL/IR-Y corpora is half the size of the other three corpora.

It can be argued quite confidently that texts included in the corpora are as representative as is practically possible, given data collection challenges, of good third-year writing from the targeted programmes. Balance is as far as possible achieved through a relatively even proportion of text contribution from each of the participant students as illustrated in Tables 5 to 8 below (numbers in the first row of each table are respectively the numbers I allocated to each student in my study (in the X corpora) and the numbers allocated to each students within the BAWE study (in the Y corpora)).

| student | 001 | 007 | 009 | 010 | 011 | 012 | 013 |
|---------------|-----|-----|-----|-----|-----|-----|-----|
| No. of essays | 3 | 2 | 6 | 6 | 5 | 5 | 5 |

Table 5: HIST-X

| student | 0019 | 0029 | 0040 | 0042 | 0144 | 0244 | 0252 | 0255 | 0314 | 0318 | 0380 | 0391 |
|---------------|------|------|------|------|------|------|------|------|------|------|------|------|
| No. of essays | 4 | 4 | 3 | 2 | 2 | 3 | 2 | 2 | 1 | 3 | 3 | 2 |

Table 6: HIST-Y

| student | 002 | 003 | 004 | 005 | 006 | 007 | 008 |
|---------------|-----|-----|-----|-----|-----|-----|-----|
| No. of essays | 3 | 6 | 5 | 5 | 6 | 2 | 6 |

Table 7: POL/IR-X

| Student | 0137 | 0234 | 0244 | 0324 | 0399 | 6180 |
|---------------|------|------|------|------|------|------|
| No. of essays | 3 | 1 | 4 | 1 | 1 | 3 |

Table 8: POL/IR-Y

3.5.2 A closer look at the corpus data - the essay questions

It would be useful to have as detailed a description as is practically possible of the essay data making up the four sub-corpora. It is beyond the scope of this project to conduct a detailed macro-level genre-focused analysis of the essay texts in each sub-corpus in terms of, for example, moves, steps or genre stages. Nevertheless, having a broad understanding of the focus and purpose of the target essays texts across sub-corpora may ultimately prove useful in making sense of results yielded from corpus-driven phraseological analysis central to the project.

It was also impractical within the time constraints of the project to closely survey each text; instead, a survey of the essay questions has been conducted, and, in three cases, where these did not provide adequate information, an analysis of the first paragraphs of the text files has been carried out (this was required for HIST-Y texts 0318f, 0318g and 0318g for which the titles were not complete essay questions).

It was my initial intention to exploit Nesi and Gardner's (2012, p. 98) six Essay genres, exposition, discussion, challenge, factorial, consequential, and commentary, as a classificatory framework for this process. However, this approach proved unsatisfactory for the purposes of building a nuanced picture of the nature of the target essay texts from their titles. Whilst the Essay genres provide a potentially useful description of different broad argumentative patterns and the genre stages through which they are realised, they proved problematic as a tool for my particular purposes. It was not always possible to link essay questions to particular Essay genres: for example, many of the questions could have been categorised either as exposition or discussion and differentiation between these options would have required a detailed reading of each text. Another problem with this approach was that, in many cases, essay titles seemed to require a combination of genre approaches; for example, the question 'To what extent can the longevity of the Franco regime be explained in terms of a series of favourable international

contexts?’ seems to be a combination of an exposition/discussion approach and a factorial approach.

Ultimately, I adopted a bottom-up method for categorising the essay questions in terms of the focus or foci which they seemed to be inviting in a response. Five broad categories of focus emerged. The first pattern of focus identified was a *focus on the specific* in terms of actors, contexts or phenomena. In essays with this focus, the questions involve analysis of particular actors, contexts, events and phenomena. In contrast, the second pattern of focus identified was a *focus on general concepts* potentially applicable across a range of specific instances.

Within both of the History sub-corpora, assignment questions in almost all cases involve focus on the specific. In contrast, a greater proportion of questions in the two Politics/International Relations sub-corpora focus on general concepts, although this preference is not as marked as the preference for specificity of focus in History. The four essay titles from each of the sub-corpora in Table 9 below provide an illustration of this difference.

| | | |
|----------|--|----------|
| HIST-X | How far did the successes of the Albigensian crusade owe to the personal achievements of Simon de Monfort? [HIST3_009_D] | specific |
| HIST-Y | Why were Chinese luxury goods of so much fascination to seventeenth and eighteenth century Europeans? [0019g] | specific |
| POL/IR-X | Is Hegel right to think that freedom is actualized through the ethical life of a modern state? [PIR3_003_Q] | general |
| POL/IR-Y | Should surfers be fed? A Critical examination of Arguments for and Against Unconditional Basic Income [0244k] | general |

Table 9: 'specific' vs. 'general' focus

A third pattern of focus identified was that of *explanation of cause-effect processes*. Language associated within this focus within assignment questions includes *why, for what reasons, what factors account for* and a range of other phrases either expressing causality or impact such as *the role of, blame for, because of* and *implications of*.

A fourth pattern of focus is *a focus on ideas*, such as ‘liberty’, *theories*, such as ‘the dialectic of lordship and bondage’ or *issues*, such as ‘capital punishment’ (N.B. ideas and theories predominate with ‘issues’ making up a much smaller proportion). In questions of this kind we frequently find phrases for expressing the explanatory power of particular ideas or theories, such as ‘*Can ... be used to understand ...?*’, ‘*What does ...tell us about ...?*’, ‘*what does it say about ...?*’ and ‘*To what extent is ... useful when analysing ...?*’.

The cause-effect focus overlaps strongly with focus on the specific, whilst the focus on ideas, theories and issues overlaps strongly with a focus on general concepts. Thus, there are more instances of the former focus to be found in both of the History sub-corpora, and more instances of the latter focus to be found in both of the Politics/International Relations sub-corpora. Table 3.9 above also serves to illustrate this difference between the History and the Politics/International Relations corpus texts.

A final pattern of focus identified (illustrated in Table 10 below) was a focus on *exploration and evaluation of the nature of phenomena*. Questions with this focus a close interrogation and discussion of what a past phenomenon was like (as in text HIST3_009_C) or is like what a current phenomenon is like (as in text 0137m).

| |
|---|
| ‘In the middle ages, magical medicine and religious medicine were essentially different examples of the same thing’. Discuss. [HIST-X: HIST3_009_C] |
| Threat of Islam or threat to Islam? Critically assess the conflict between Islam and the West. [POL/IR-Y: 0137m] |

Table 10: Exploration and evaluation of the nature of phenomena

Appendix 2 contains a full breakdown of how each essay text can be categorised. These categories are not mutually exclusive. For example, the following title (in Table 11) from the HIST -RHUL sub-corpus contains both focus on specificity (Oppenheimer) and focus on ideas, theories or issues (in this case a moral issue).

With reference to Oppenheimer, reflect on the moral dilemma of scientists in disseminating research on the one hand, and developing weapons of mass destruction on the other. [HIST-X: HIST3_012_B]

Table 11: Focus on specificity & ideas/theories/issues

Similarly, the following question (in Table 12) from POL/IR-Y sub-corpus contains both focus on ideas, theories or issues (in this case on theory) as well as exploration/evaluation of the nature of phenomena (in this case, the nature of EU policymaking).

To what extent are the Grand Theories of Integration useful tools for analysing European policymaking? [POL/IR-Y: 0244j]

Table 12: Focus on ideas & exploration/evaluation of the nature of phenomena

Table 13 below shows a proportional breakdown of the foci within essay questions across disciplinary/institutional sub-corpora. Any given essay can be placed in more than one category so the percentages shown are correct despite not adding up to 100% (it needs also to be noted that four repetitions of the question ‘Does Ranke deserve the title ‘father of scientific history’?’ occur in the HIST-Y sub-corpus. This question has been judged to carry both focus on cause-effect - Ranke’s impact on the discipline of History - and on evaluation of nature of phenomena - the nature of Ranke’s approach - so contribute to percentages in both categories and very slightly skew proportions for HIST-Y).

| Focus on | HIST-X | HIST-Y | POI/IR-X | POL/IR-Y |
|--|------------|------------|------------|------------|
| <i>the specific</i> (actors, contexts or phenomena) | 94% | 83% | 39% | 31% |
| <i>general concepts</i> applicable across a range of instances | 6% | 17% | 61% | 69% |
| <i>explanation of cause-effect processes</i> | 67% | 56% | 27% | 23% |
| <i>ideas, theories or issues</i> | 16% | 20% | 54% | 54% |
| <i>evaluation of the nature of phenomena.</i> | 16% | 53% | 15% | 23% |

Table 13: Proportions of types of essay focus across sub-corpora

This represents a very broad-brush survey of the essay questions as, within the practical time constraints of the project, there was no room for more fine-grained attention to this task. Nonetheless, this survey has yielded a useful indication of disciplinary differences in terms of focus between essay texts in History and Politics/International Relations. History favours a focus on the specific as well as cause-effect. Politics/International Relations, by contrast, favours a focus on general widely applicable concepts and on ideas, theories and issues.

3.6. Keyword analyses

3.6.1. Choice of reference corpus: the BNC Sampler - written component

Scott and Tribble (2006, pp. 59-64) illustrate the impact of the choice of reference corpus on the content of a keyword list generated by conducting a keyword analysis of a research corpus consisting of the text of Shakespeare's *Romeo and Juliet* with two different reference corpora, all of Shakespeare's plays and the British National Corpus respectively. The former throws up proper noun names of specific characters within *Romeo and Juliet* as well as words reflecting central themes such as *love*, *death* and *poison*. The latter also throws up words reflecting specific

characters and central themes and a set of words not present in the former list which reflect the specific nature of Shakespearean language such as *thou*, *thy*, *O* and *hath*. Whilst underlining the important impact of reference corpus choice in this way, they argue this demonstration at the same time shows that ‘while the choice of reference corpus is important, above a certain size, the procedure throws up a robust core of KWs whichever the reference corpus used’ (ibid., p. 64). In a later article, Scott argues again for the robustness of the keywords procedure maintaining that keywords ‘identified even by an obviously absurd [reference corpus] can be plausible indicators of aboutness’ (2010, p. 51). However, he also points out that, in terms of content of reference corpus, genre and whether texts are spoken or written have a significant impact (ibid.).

As discussed above in section 2.3.5. of the Literature Review, in the context of usefulness of keywords to discourse analysis, Baker (2004) highlights the danger of overemphasis of difference if two research corpora are compared with each other in generating keyword lists. Taylor (2013) also takes up the issue of too much focus on difference in corpus work, arguing for the need for greater focus on similarity for two reasons; firstly, she argues that ‘by focusing on difference, we effectively create a ‘blind spot’; this means that, rather than aiming for a 360-degree perspective on our data, we are actually starting out with the goal of achieving only 180-degree visualisation’ (p. 83); secondly, she argues that ‘by setting out to look at difference, the analyst is likely to find and report on difference’ which creates ‘a significant threat to the balance of analysis’ (ibid.). She cites Baker’s idea of the ‘bottom drawer syndrome’ in which researchers who find similarity tend to file rather than publish such findings leading to a picture in the published research comparing a particular set of discourses or language types of a greater degree of difference than actually exists (Baker, 2010, p. 83, cited in Taylor, 2013, p. 83).

Taking into account Baker’s and Taylor’s arguments, I have aimed in my project for a research design which is able to account for both similarity and difference. For this reason, rather than compare research corpora against each other, which would risk exaggerating disciplinary and/or institutional differences at expense of possible

similarities, I have chosen a larger, more general corpus, the BNC Written Sampler (discussed below), as reference corpus in keyword analyses of all four sub-corpora. Using the keyword lists generated in this way, I have identified both similarities and differences in terms of keywords present across all four corpora. Taking into account also the impact of whether texts in a reference corpus are spoken or written (Scott and Tribble, 2006), I have opted for a reference corpus made up solely of written texts. This is because I am not so much concerned with highlighting features which single out the texts in my research corpora as written as opposed to spoken texts, but rather, with highlighting features by which those texts can be compared or contrasted as written texts belonging to specific disciplinary discourse communities. For this reason also, the BNC Written Sampler is a useful reference corpus for my research purposes.

The BNC was chosen because, in comparison to the 'specialised' nature of the four sub-corpora under examination, which only include written texts in the essay genre from the academic domain, it is a 'general' corpus the purpose of which is the study of modern British English as a whole. The BNC is 'a well-known general corpus' (McEnery et al., p. 59) consisting of 4,124 texts contributing to a total of 100,106,008 words of modern British English, 90% of which is written, 'samples from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, as well as school and university essays' (ibid.), and 10% of which is spoken, 863 transcripts of informal conversation including a balance demographically in terms of social class, region and age, and a range of different contexts from 'formal business and government meetings to radio shows and phone ins' (pp. 59-60).

The BNC Sampler, created at Lancaster University, is a 2 million word sub-corpus of the BNC created in order to manually check and correct word class tagging and also to create a corpus in which the balance between the written and spoken elements were evenly balanced with an approximate 50%-50% division (UCREL, Lancaster University, 1998). It contains 'a wide and balanced sampling of texts from the BNC,

so as to maintain the general text types and the proportions of general text types (apart from the unequal written/spoken division) of the BNC as a whole' (ibid.). The written portion of the BNC Sampler can be drawn on as a reference corpus in *Wmatrix*, the programme I used to create my keyword lists (discussed below). The written portion of the BNC Sampler's use as a reference corpus in comparison to the History and Politics/IR corpora from both institutions will show words that are unusually frequent in the written texts in these corpora in comparison to their frequency in 'general' usage in written texts beyond the academic domain. Differences between the keyword lists generated for the four corpora may indicate disciplinary and/or institutional variation.

There are arguments against the BNC's use as a reference corpus in circumstances where the time period in which it was developed, compiled from the 1980s to 1993, could significantly skew the nature of the keyword lists generated; changes, for example in society, politics or technology, since this time have been likely to impact language usage in terms of the salience of particular themes and consequently the frequency of particular content lexis or proper names in spoken and written texts (e.g. Johnson and Ensslin, 2006). However, although, the BNC is arguably somewhat 'dated' at this point, this is likely to have very little or no significant impact on the occurrence of closed-class grammatical words, which are to be the focus of my study (discussed below), in keyword lists generated, and therefore the BNC's use can be justified within the context of my project.

3.6.2. Keyness analysis procedure and identification of items for analysis

I used *Wmatrix* to run a keyness analysis for each of the four corpora using as the reference corpus the 'BNC Sampler Written' available for selection as part of this software tool. I filtered the resulting list to include only overused words (positively key), to exclude 'multi-word items' from being calculated automatically (these items change the frequency/keyness of individual items that make up a multi-word grouping), and with the log likelihood (LL) cut-off of 6.63 and the frequency cut-off

of 5 suggested within the programme¹. I then manually identified all 'closed-class grammatical' items in all four lists. In identifying grammatical words, I took into account the loose definitions for these given by both Scott and Tribble (2006) and Groom (2007). Scott and Tribble (2006, p. 98) cite the definition by Richards et al of 'function words' as 'those which "have little meaning on their own but which show grammatical relationships in and between sentences" (Richards et al. 1985:61)'. Groom define (2007, p. 52) closed-class words as 'determiners, prepositions, pronouns and conjunctions, plus other high frequency words with a largely grammatical function (auxiliary and modal verbs, adverbs of negation, and so on)'.

From this data I collated the following lists: closed-class grammatical items that were key across all four corpora (in Table 14 below); closed-class grammatical items which were key in one discipline across both institutions (i.e. either key only in the two History corpora or only in the two Politics/PIR corpora) (Table 15); closed-class grammatical items key only in one institution (i.e. key in both History and Politics/PIR, but at only one institution) (Table 16); items key in three out of the four sub-corpora (Table 17); items key in one corpus (Table 18); and, items key in a random pairing of sub-corpora (i.e. different disciplines at different institutions)(Table 19).

Noting Scott and Tribble's argument that 'the order of keywords is not intrinsically trustworthy' (2006, p. 50), and noticing also that the log likelihood figure changes in relation to the size of the corpus being analysed, I decided to base the sequence in my lists on the stable figure of the percentage frequency of each lexical item within the corpora. In Tables 14 to 19, three pieces of data are listed for each keyword in each sub-corpora: the position that it appears on the keyword list for the sub-corpus, the number of times it occurs within the particular sub-corpus, and its percentage frequency within the sub-corpus.

¹ Wordsmith cuts off at 500 words – LL and frequency cut-offs more useful – '500' = arbitrary – the 500th word in two different corpora may have a different level of LL/frequency

When selecting potential items for further analysis I chose as a cut-off point at least 300 instances of each word in each corpus. A cut-off of 150 was allowed for the POL/IR-Y corpus because of its smaller size, roughly half the length of the other three corpora. This number is slightly arbitrary, but has been chosen because 300 is the size of random samples (discussed as part of explanation of the analytical procedure for further analysis of the keywords in section 3.7 below)

Tables 14 to 19 below show the keyword analyses, filtered as described above. All closed class key items are present, but those which were eliminated due to low numerical frequency are presented struck through. Blank squares represent lack of positive keyness for a particular item in a particular sub-corpus. In Table 14 it is interesting to note that on the whole frequencies for each of the items are very similar, with the exception of *this* and *within* which are both noticeably higher in frequency in the POL/IR-X sub-corpus. It is also interesting to note that the frequencies of the first five items, *of*, *and*, *that*, *as* and *this*, when added together constitute roughly 10% of all text in each sub-corpus. Of these five items, position on the keyword list for *of*, *that*, *as* and *this* tends to correspond to level of frequency within the text. The keyword *and* is a slight anomaly in that although in all four sub-corpora it represents the second most frequently occurring item, it occurs much further down the keyword list.

| | HIST-X | | | HIST-Y | | | POL/IR-X | | | POL/IR-Y | | |
|---------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|
| | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % |
| of | 1 | 3641 | 4.82 | 4 | 4257 | 4.50 | 5 | 4053 | 4.78 | 35 | 1943 | 4.11 |
| and | 491 | 2383 | 3.16 | 207 | 3080 | 3.26 | 282 | 2775 | 3.28 | 537 | 1489 | 3.15 |
| that | 4 | 1142 | 1.51 | 11 | 1334 | 1.41 | 19 | 1255 | 1.48 | 41 | 648 | 1.37 |
| As | 15 | 732 | 0.97 | 23 | 850 | 0.90 | 12 | 943 | 1.11 | 40 | 448 | 0.95 |
| this | 24 | 587 | 0.78 | 60 | 645 | 0.68 | 13 | 815 | 0.96 | 163 | 297 | 0.63 |
| between | 70 | 170 | 0.23 | 148 | 172 | 0.18 | 102 | 188 | 0.22 | 314 | 81 | 0.17 |
| within | 91 | 98 | 0.13 | 326 | 83 | 0.09 | 30 | 182 | 0.21 | 200 | 54 | 0.11 |
| itself | 305 | 41 | 0.05 | 526 | 38 | 0.04 | 145 | 64 | 0.08 | 228 | 31 | 0.07 |

Table 14: Closed-class grammatical items key in each of the four corpora

In Table 15, which shows items with discipline-specific keyness, the frequencies are also similar for each item in both disciplinary corpora in which it is key. Considering the nature of the subject areas involved, it is not really surprising that present tense

verb forms, *is*, *can* and *does*, appear in the Politics/International Relations sub-corpora while past tense verb forms, *was* and *were*, appear in the History sub-corpora. The tendency, discussed in section 3.5.2 above, for focus to be on specific phenomena in the History texts compared to the tendency in the Politics/International Relations texts to deal with generalised phenomena may go some way towards explaining the presence of *his* and *their* as key in History, to talk about specific human actors, and the presence of the indefinite article *an* in Politics/International Relations, to talk about non-specific phenomena.

| | HIST-X | | | HIST-Y | | | POL/IR-X | | | POL/IR-Y | | |
|----------|------------------|-----------------------|------------------------|------------------|-----------------------|------------------------|------------------|-----------------------|------------------------|------------------|-----------------------|------------------------|
| | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % |
| is | | | | | | | 25 | 1390 | 1.64 | 47 | 724 | 1.53 |
| an | | | | | | | 69 | 444 | 0.52 | 134 | 218 | 0.46 |
| can | | | | | | | 92 | 306 | 0.36 | 53 | 153 | 0.32 |
| its | | | | | | | 181 | 299 | 0.35 | 380 | 149 | 0.32 |
| does | | | | | | | 184 | 121 | 0.14 | 473 | 53 | 0.11 |
| was | 16 | 1036 | 1.37 | 30 | 1213 | 1.28 | | | | | | |
| were | 29 | 471 | 0.62 | 33 | 560 | 0.59 | | | | | | |
| his | 69 | 475 | 0.63 | 112 | 559 | 0.59 | | | | | | |
| their | 125 | 305 | 0.40 | 165 | 365 | 0.39 | | | | | | |
| although | 363 | 65 | 0.09 | 567 | 73 | 0.07 | | | | | | |
| himself | 406 | 42 | 0.06 | 606 | 45 | 0.05 | | | | | | |
| toward | 248 | 12 | 0.02 | 415 | 10 | 0.01 | | | | | | |

Table 15: Closed-class grammatical items key in only one discipline

In Table 16 showing items key in one institution, again, frequencies across sub-corpora are very similar. However, only one item, *by*, which is only key at Institution X, occurs frequently enough to be deemed potentially of value for further analysis.

| | HIST-X | | | HIST-Y | | | POL/IR-X | | | POL/IR-Y | | |
|------------|------------------|-----------------------|-------------|------------------|-----------------------|------------------------|------------------|-----------------------|-------------|------------------|-----------------------|------------------------|
| | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % |
| by | 447 | 529 | 0.70 | | | | 435 | 604 | 0.71 | | | |
| more | | | | 280 | 263 | 0.28 | | | | 148 | 161 | 0.34 |
| these | 270 | 148 | 0.20 | | | | 309 | 166 | 0.20 | | | |
| through | 207 | 106 | 0.14 | | | | 266 | 117 | 0.14 | | | |
| themselves | 317 | 38 | 0.05 | | | | 501 | 38 | 0.04 | | | |
| amongst | 555 | 19 | 0.02 | | | | 415 | 13 | 0.02 | | | |
| whereas | | | | 626 | 11 | 0.01 | | | | 355 | 9 | 0.02 |

Table 16: Closed-class grammatical items key at only one institution across both disciplines

In Table 17 showing words which are key in three out of the four sub-corpora, again, frequency is remarkably similar across the three corpora in which each item is key. Only three items, *the*, *in*, and *which* occur frequently enough to be deemed potentially useful for further analysis.

| | HIST-X | | | HIST-Y | | | POL/IR-X | | | POL/IR-Y | | |
|---------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|
| | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % |
| the | 6 | 6316 | 8.37 | 26 | 7490 | 7.92 | 34 | 6799 | 8.02 | | | |
| in | 249 | 1615 | 2.14 | 175 | 2061 | 2.18 | | | | 363 | 1000 | 2.12 |
| which | 283 | 377 | 0.50 | | | | 293 | 429 | 0.51 | 172 | 264 | 0.56 |
| both? | 446 | 87 | 0.12 | 632 | 102 | 0.10 | 634 | 92 | 0.11 | | | |
| against | 271 | 84 | 0.11 | | | | 669 | 76 | 0.09 | 408 | 49 | 0.10 |
| often? | 353 | 50 | 0.07 | 494 | 56 | 0.06 | | | | 113 | 51 | 0.11 |
| towards | 573 | 39 | 0.05 | 700 | 45 | 0.05 | 285 | 59 | 0.07 | | | |
| because | | | | 621 | 84 | 0.08 | 690 | 74 | 0.09 | 460 | 43 | 0.10 |

Table 17: Closed-class grammatical items key in three corpora

Tables 18 and 19 respectively show items key in only one sub-corpus and items key in a random disciplinary/institutional pairing of sub-corpora, with only *had* in Table 3.18 and *to* and *not* in Table 3.19 occurring frequently enough to be deemed worth potential further investigation.

| | HIST-X | | | HIST-Y | | | POL/IR-X | | | POL/IR-Y | | |
|------------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|
| | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % |
| had | 548 | 321 | 0.43 | | | | | | | | | |
| would | 497 | 239 | 0.32 | | | | | | | | | |
| who | 287 | 230 | 0.30 | | | | | | | | | |
| how | 336 | 98 | 0.13 | | | | | | | | | |
| throughout | 293 | 35 | 0.05 | | | | | | | | | |
| amidst | 565 | 5 | 0.01 | | | | | | | | | |
| upon | | | | 399 | 52 | 0.05 | | | | | | |
| or | | | | | | | 708 | 338 | 0.40 | | | |
| why | | | | | | | 469 | 68 | 0.08 | | | |
| beyond | | | | | | | 785 | 22 | 0.03 | | | |
| it | | | | | | | | | | 519 | 460 | 0.97 |
| are | | | | | | | | | | 219 | 327 | 0.69 |
| has | | | | | | | | | | 69 | 244 | 0.52 |
| may | | | | | | | | | | 354 | 84 | 0.18 |
| while | | | | | | | | | | 227 | 57 | 0.12 |
| above | | | | | | | | | | 486 | 31 | 0.07 |

Table 18: Closed-class grammatical items key in one corpus

| | HIST-X | | | HIST-Y | | | POL/IR-X | | | POL/IR-Y | | |
|-----|------------------|-----------------------|-------------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|------------------|-----------------------|-------------|
| | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % | Position on list | Number of occurrences | Frequency % |
| to | 126 | 2132 | 2.82 | | | | | | | 478 | 1244 | 2.63 |
| not | | | | 391 | 524 | 0.53 | 130 | 539 | 0.64 | | | |

Table 19: Closed-class grammatical items key in a random pairing of the four corpora

As will be explained in Chapter 4, due to the time-consuming and highly detailed nature of the analytical procedure followed in the project, ultimately only keywords from Table 3.14, key across all four sub-corpora, were the subject of further analysis within this project.

3.7. Approach/procedure for further analysis of identified/selected keywords

In undertaking a bottom-up, semantically-focused analysis of the phraseologies around the key grammatical items identified in the four disciplinary/institutional corpora, this study broadly follows the approach devised by Groom (2007) thereby accepting his invitation to further test the methodology (p. 89). Groom ventures that the claims generated by the results of his analysis ‘are both plausible in themselves and amenable to further testing (if not exact replication) by subsequent research’ (ibid., p. 89). The current study is not an exact replication of Groom’s, but will be testing the efficacy of the approach when applied to a different genre family, the student Essay, using the much smaller corpora that are practically possible to assemble for this genre. While the broad purpose and approach mirrors Groom’s, the study has departed from exact replication in some aspects of the procedure. It is therefore useful at this point to outline Groom’s procedure in some detail before explaining and justifying how the current study both maps with and diverges from it.

3.7.1 Groom's (2007) approach

Groom models the analytical procedure used in his study by providing a detailed description of his treatment of one particular grammatical keyword in one disciplinary corpus. He takes the reader on a step by step journey through his analysis of random concordance-line samples for the keyword ***of*** within his HistArt corpus of Research Articles from History (2007, pp. 81-100).

He analyses three 100-line random samples from HistArt in turn, in each case first analysing and grouping lines according to 'structural properties' before going on to analyse them from a 'semantic perspective' (p. 82). Any 'semantic sequence' (explained below) occurring at least twice in a 100-line sample is given a label, whilst sequences that occur only once are relegated to 'Other'.

With each new 100-line sample there is the possibility of the addition of a new semantic sequence category to the taxonomy. Figure 2 below shows the table from Groom's study recording the structural analyses across the three samples (2007, p. 99). The COBUILD metalanguage is used by Groom to code grammatical categories, with bold lower case letters used to denote particular word classes and bold italics to denote particular word forms. Thus, for example, ***n of n*** denotes post-modified noun-phrases.

| Sequences | Examples | Sample 1(%) | Sample 2(%) | Sample 3(%) |
|-------------------------------|---|-------------|-------------|-------------|
| n of n | this logic prevailed until <u>the second feminist wave of the 1970s.</u> | 90 | 92 | 87 |
| prep n | they were eventually forced to seek help from local people, particularly from farmers <u>in the vicinity of camps.</u> | 3 | 5 | 9 |
| adj of n | most of us are <u>aware of the rules of evidence</u> | 1 | 2 | 2 |
| v n out of n | the laws acted to <u>take weapons out of men's hands</u> | 2 | 1 | 0 |
| v of n | Bainsville ... <u>wrote of this recurrent dilemma of French foreign policy</u> | 1 | 0 | 2 |
| fixed phrase | they facilitated the entry of respectable women into what <u>one turn-of-the-century</u> writer termed the "Night Side of London" | 1 | 0 | 0 |
| the adj-superl of pl-n | the dubbing of a knight is <u>the most familiar of the new ceremonies</u> | 1 | 0 | 0 |
| v n of n | Chulaki <u>accused them both of cosmopolitanism</u> | 1 | 0 | 0 |

Table 4.10: Structural analysis of **n of n** sequences in third sample

Figure 2: Table from Groom, 2007, p. 99

Figure 3 below shows a table of the subsequent semantic analyses of the three samples. For the 'semantic sequences' identified, bold small caps are used for the 'semantic elements' that contribute to the sequence, the keyword is italicised and in bold, and a '+' is used to show how elements and the keyword combine. Thus, for example, 'the basic tenets of Marxist theory' is grouped in the semantic sequence category **PROPERTY + *of* + PHENOMENON**.

| Semantic sequence | Example | Sample 1 (%) | Sample 1 (%) | Sample 1 (%) | Average |
|--|---|--------------|--------------|--------------|---------|
| PROPERTY + <i>of</i> + PHENOMENON | the basic tenets of Marxist theory | 14 | 13 | 16 | 14.333 |
| CONCEPTUALISATION + <i>of</i> + PHENOMENON | a tidal wave of conservative loyalism | 7 | 15 | 15 | 12.333 |
| PROCESS + <i>of</i> + OBJECT | control of female sexuality | 11 | 17 | 15 | 14.333 |
| PART + <i>of</i> + WHOLE | the east wall of that tiny church | 3 | 5 | 8 | 5.333 |
| QUANTITY + <i>of</i> + PHENOMENON | Hundreds of thousands of shares | 7 | 8 | 7 | 7.333 |
| PROCESS + <i>of</i> + ACTOR | the death of Henry II | 6 | 7 | 6 | 6.333 |
| QUALITY + <i>of</i> + PHENOMENON | the insignificance of the English | 3 | 7 | 5 | 5.000 |
| AUTHORITY + <i>of</i> + DOMAIN | the third Duke of Northumberland | 9 | 4 | 4 | 5.666 |
| GROUP + <i>of</i> + MEMBERS | a transatlantic 'community of ideals and interests' | 4 | 2 | 3 | 3.000 |
| ACTOR + <i>of</i> + OBJECT | the worker of injustice | 5 | 1 | 2 | 2.666 |
| PHENOMENON + <i>of</i> + TIME | the sweeping utopian projects of the postwar period | 4 | 3 | 2 | 3.000 |
| TEXT + <i>of</i> + CONTENT | Journal of Imperial Commonwealth History | 9 | 2 | 2 | 4.333 |
| TIME + <i>of</i> + QUALITY | six successive years of disappointment | 0 | 2 | 2 | 1.333 |
| INSTITUTION + <i>of</i> + LOCATION | the kingdom of Germany | 0 | 2 | 0 | 0.666 |
| NAME + <i>of</i> + LOCATION | Hildegard of Bingen | 2 | 3 | 0 | 1.666 |
| RESULT + <i>of</i> + PROCESS | the result of civil war | 3 | 0 | 0 | 1.000 |
| RELATION + <i>of</i> + INDIVIDUAL | widow of the first leader | 2 | 0 | 0 | 0.666 |
| Other | the feast of St. John the Baptist | 1 | 0 | 0 | 0.333 |
| TOTALS | | 90 | 92 | 87 | 89.327 |

Table 4:11: Semantic analysis of *n of n* sequences in third sample

Figure 3: Table from Groom, 2007, p. 100

The results of this analysis of three consecutive 100-line random samples for *of* are used by Groom to justify his choice in analysis of all subsequent grammatical keywords to examine *only one* 100-line sample for HistArt and for LitArt (Literature) respectively. He argues '100 lines of data are a sufficient basis for making statements of proportion and/or frequency' due to the fact that the overall conclusions drawn from analysis of the three 100-line samples for *of* would have been exactly the same had only one 100-line sample been analysed (2007, p. 101).

Groom further justifies this choice using the following logic: the fact there is such strong continuity across the three 100-line samples for a highly frequent word such as *of* suggests that keywords which are less frequent will exhibit the same level of continuity between samples (ibid.).

This assertion of Groom's needs to be treated with a level of caution. Firstly, the consistent behaviour of one salient grammatical keyword across a sequence of random samples can only confidently be argued to be evidence of regular behaviour for *that particular* grammatical word; this regularity may not necessarily transfer to behaviour of another grammatical keyword. Furthermore, even if we were to accept the argument that a salient grammatical item will behave in a consistent manner, there still remains the chance that a particular random sample will be irregular or unrepresentative in some way. It would thus seem unwise not to at the very least test further via second samples any differences between disciplinary corpora which appear in the first samples.

Although not present in his demonstration analysis of *of* where we only find semantic sequences, Groom makes a useful distinction between 'semantic sequences' and 'semantic motifs' in his study (2007, p. 102). The former is a phraseological pattern sharing both meaning and patterning of semantic elements, such as phrases which, for example, share the **PROPERTY + of + PHENOMENON** pattern. The latter is an umbrella term coined by Groom to describe 'a group of semantic sequences or other phraseological items which are grouped together insofar as they share a similar broad meaning' which are expressing 'similar meanings in slightly different ways' (ibid.). Semantic motifs are indicated with single quotation marks. For example, the semantic sequences for the keyword **among** are grouped in Groom's study under the two semantic motifs, 'phenomenon in context' and 'member of group'.

A feature of Groom's approach which warrants further attention is the fact that with each grammatical keyword a structural analysis is conducted before the semantic analysis. Groom explains that whilst the central focus of his study is

‘semantic features and other phraseological features’, he has ‘found it useful to perform an initial structural analysis on each data set so as to break it down into more manageable parts, as suggested by Sinclair (2003)’ (2007, p. 82). He acknowledges that this choice is open to question given that ‘semantic sequences are not tied to particular grammatical structures’. He justifies it along personal and practical lines: ‘I found it easier to start by dividing the data up into structures first, then searching each of these structures for repeated semantic sequences, and then finally comparing different sets of sequence/structure complexes and amalgamating any where the semantic sequence elements were the same’ (ibid., p. 84).

The interplay between semantic and grammatical patterns is attended to within Groom’s analysis. In discussion of semantic sequences, the relative strength of associations between semantic and grammatical/structural patterns are highlighted. For example, in his analysis of **among** Groom describes ‘phenomenon in context’ sequences as ‘very strongly associated with the nominal structure **n among pl-n**’ (p. 104) and as such with ‘phrasal structures’, while ‘member of group’ sequences are strongly (although ... not exclusively) associated with clausal structures’ (p. 109).

Where it is deemed useful, terminology from existing grammatical models is deployed by Groom as part of his semantic labelling. For example, some of the labels for semantic elements are drawn or adapted from systemic functional linguistics: ‘**PROCESS**’ is ‘used in the functional terms of Halliday (1994)’ (p. 85); the broader term ‘**OBJECT**’ is chosen to describe the phenomenon affected by the ‘**PROCESS**’ as Groom maintains that ‘for the purposes of the present thesis at least I have not deemed it necessary to adopt Halliday’s rather complex nomenclature for describing things affected by processes (i.e. “Range”, “Goal”, etc.)’ (p. 85). For labelling semantic elements of semantic sequences grouped under the semantic motif ‘phenomenon in context’, glosses from Pattern Grammar (Francis et al., 1998, pp. 128-130) such as ‘**FEELING OR ATTITUDE**’, ‘**ESTEEM**’ and ‘**CONFLICT**’ are deployed (Groom, 2007, pp104-105).

Grammatical features are nonetheless very much subservient to the central semantic aim of Groom's approach. Indeed, they are supposed to do no more than fulfil a very much practical preliminary function in his analyses. However, at places, attention to grammatical features of phraseologies detracts from and dilutes Groom's semantic analysis. Three salient grammatical items in Groom's study raise problems in this respect.

Firstly, in his analysis of the grammatical keyword **and** (2007, pp. 114-122), Groom's semantic categorisation is divided under the umbrella structural categories of 'phrasal sequences' and 'clausal sequences'. Whilst semantic groupings may fall entirely along these lines, the final semantic analysis arguably should not retain such explicitly grammatical labels.

Secondly, in his analysis of the fixed phrase **as well as** (p. 123) (part of analysis of the keyword **as**) he uses the systemic function linguistic text-organisational concepts of **GIVEN INFORMATION** and **NEW INFORMATION** as labels for elements in a semantic sequence, i.e. '**GIVEN INFORMATION** + **as well as** + **NEW INFORMATION**'. Although SFL labels, especially experiential ones such as **PROCESS**, can sometimes be usefully appropriated for semantic labelling purposes, text organisational labels such as **GIVEN INFORMATION** and **NEW INFORMATION** are departing too far from the semantic spirit of the procedure.

Finally, in his analysis of **these** (pp. 183-185) Groom departs almost entirely from analysis of the semantic phraseologies in which this keyword occurs and, instead, analyses and categorises it according its 'text cohesive features' drawing on work by Sinclair (2004), Francis (1985, 1994) and McCarthy (1994); this involves identification of whether **these** is referring to ideas in the text preceding (i.e. Sinclair's *encapsulation*) or following it (i.e. Sinclair's *prospection*) and whether they involve summary, either neutral or evaluative, or repetition of the ideas being referred to. The resulting table is labelled 'Grammatical and semantic roles of **these** in HistArt and LitArt'; thus, grammatical analysis has moved very much beyond its purely practical role in this particular instance (N.B. Groom's analysis of **these** will

be discussed in more detail in Chapter 4 as part of a contrast between his approach and the approach taken to the keyword **this** within the current study).

3.7.2 The approach used in this study

My analysis of salient grammatical keywords maps exactly with the approach devised by Groom in many important ways. Random sampling is conducted and the phraseologies of each keyword are analysed and categorised according to their semantic properties. Groom's concepts 'semantic sequence' and 'semantic motif' are adopted as are the majority of his labelling conventions and his cut-off point of 2% for the inclusion or exclusion of a semantic grouping.

However, my approach differs to Groom's in some important aspects and Table 20 gives an overview of the key procedural differences. The remainder of this section explains these differences and also provides further detailed explanation of the procedure I have followed.

| Groom (2007) | This study |
|--|---|
| 100-line random samples | 300-line random samples |
| <ol style="list-style-type: none"> 1. Analysis of one sample from HistArt 2. Followed by analysis of a second sample from LitArt | <ol style="list-style-type: none"> 1. Analysis of one sample from a <i>pooled corpus</i> containing all four disciplinary/institutional sub-corpora 2. Followed by analysis of a sample from each of the four sub-corpora 3. Analysis of second samples where necessary to verify any differences identified in analysis of the first set of samples |
| Concordance lines grouped first according to structural properties, then regrouped according to semantic properties. | Concordance lines grouped <i>according to semantic properties only</i> . |

Table 20: Comparison/Groom of key features of Groom's approach (2007) and the approach taken in this study

For all words after **of** in his study Groom analysed one 100-line random sample from HistArt followed by one 100-line sample from LitArt. In contrast, for every word in this study the choice was made to start with analysis of a 300-line sample from a 'pooled' corpus which contained all texts from the four sub-corpora before then moving on to analysis of a 300-line random sample from each individual

disciplinary/institutional sub-corpora. Lines in the four sub-corpora would be categorised according to semantic groupings identified in analysis of the pooled corpus with any new semantic categories identified in the sub-corpora also added at this stage.

The rationale for starting with this pooled corpus was both to establish a basis for comparison of the four sub-corpora and to ensure this initial identification of semantic groupings within the phraseologies of the target keyword would not be subject to the risk of bias towards the characteristics of one particular discipline within one particular institution.

The rationale for using 300-line rather than 100-line random samples was that the corpora in this study are much smaller in size than Groom's meaning there is greater risk that a particular sample may not accurately represent its corpus as a whole. This risk can be somewhat reduced by using a larger random sample size, 300 lines being the largest that could be realistically coped with for manual analysis.

To account for the risk of a non-representative sample, analysis of second random 300-line samples was carried out for any finding of difference in frequency of a category across the first four disciplinary/institutional sub-corpora samples. The analysis of the second samples was targeted, only identifying and counting lines which fall into semantic groupings where differences were observed in the first samples. Considering the extremely painstaking and time-consuming nature of exhaustively categorising an entire sample, this can be seen as an efficient use of further sampling to improve the strength and plausibility of any arguments ultimately made about continuity or difference.

For practical reasons, Groom chose to group lines from his samples first according to their structural properties before starting the process of identifying semantic sequences. As discussed above, this arguably led, in some cases, to the semantic aim being diluted by or blended with analysis of grammatical properties. To avoid this potential problem, the current study dispenses entirely with the initial

structural analysis and proceeds directly to the main purpose of the process, that of grouping lines by *semantic* sequence and/or motif. In the writing up of analysis for each keyword, groupings are discussed entirely from a semantic perspective, although in line with Groom (2007) where interplay between semantic and grammatical pattern is observed this is acknowledged and discussed.

The process of semantically grouping phraseologies in a 300-line sample is somewhat messy, very time-consuming, and iterative. Analysis of each keyword requires an initial rough grouping followed by repeated phases of regrouping in the process of fine-tuning and firming up final semantic categories. To aid the grouping and regrouping process involved in semantic categorisation, WordSmith random samples were converted to Excel and the Data Validation and Sort functions used to aid in reorganisation and grouping of the lines. Data Validation created a set of numeric category options in a column to the right of the concordance lines. Lines could thus be assigned a numeric category and then the Sort function could be used to reorganise the lines numerically. This facilitated easy and quick examination of emerging groupings and the further revision and refinement of these (some of the initial groupings tended to naturally fall along grammatical lines, but grammatical organisation was never consciously aimed for).

As far as possible, a purely bottom-up approach has been attempted throughout in the formulation of semantic categories for each of the target grammatical keywords. However, the close reading of Groom's Method and Analysis sections which was entailed in gaining a proper understanding of his methodology in order to adopt it inevitably means the current study has been influenced in some places by Groom's category distinctions and labels. This is the case particularly for the keyword **of** used by Groom to provide a detailed demonstration of his procedure.

To dilute this potential influence, as much distance as possible has been put between the reading of related parts of Groom's thesis and the analytical phase of the current study. However, it should be noted that in some instances labels used by Groom such as **PHENOMENON**, **PROCESS**, **PROPERTY** and **OBJECT** have been consciously

adopted/appropriated in this study when deemed to be useful. As discussed above, the majority of his conventions for labelling of both semantic sequences and grammatical patterns have also been adopted. More details with regards to the labelling conventions for this study will now be given at the beginning of Chapter 4.

Chapter 4: Results

Table 21 provides a reminder of the salient results of the keyword analyses carried out and discussed within Chapter 3. It lists all words occurring at least 300 times that are key across all four sub-corpora, key across the two sub-corpora of one discipline, or key across the two sub-corpora for one particular institution (only occurring at Institution X).

| Key in all 4 sub-corpora | Key in History | Key in Politics/International Relations | Key at Institution X |
|---------------------------------|-----------------------------|---|----------------------|
| of and that as this | was were his their | is an can | by |

Table 21: Keywords occurring more than 300 times (or 150 times in POL/IR-Y)

4.1 Selection of items for further analysis

Because of the extremely time-consuming and detailed nature of the analytical procedure adopted for the further analysis of keywords, it has only been possible within this project to analyse the five words in the first column of Table 4.1, those that are key across all four sub-corpora. In Groom's (2007) research, a larger number of items, nineteen words - identified as key across both research corpora in his study - were the subject of further analysis, which for each word involved analysis of one 100-line concordance sample from each corpus. The smaller number of items subject to further analysis in this project can be justified, firstly, by the fact that there are four as opposed to two research corpora, secondly, by the fact the decision was made to start with analysis of a pooled sample made up of lines from all four corpora meaning that, in fact, five samples are analysed for each keyword, and, finally, by the fact that based on the smaller size of the corpora it was decided that random concordance samples to be 300-lines rather than 100-lines long.

However, with the EAP motivation of this project in mind, the narrow focus on five key grammatical items can be justified in terms of the coverage it affords of the student texts. Bruce (2011, p. 6) argues that the extent to which a feature or set of features identified through linguistic research can be argued to be pedagogically useful 'rests on the extent to which these researched features of academic texts, of themselves, can be said to operationalize the wider phenomenon of academic subject discourses realised in texts'. In other words, if a linguistic feature has been proven to be disciplinarily distinct but nonetheless occurs relatively infrequently within any particular text, then it is still of limited use to the teaching of academic writing in the target discipline.

As has already been pointed out, in each disciplinary/institutional sub-corpus within the current study, the five target grammatical keywords, *of*, *and*, *that*, *as*, and *this* alone make up approximately 10% of all text. This means in an average essay of, say, 2,000 words, we would expect roughly 200 words to be these five keywords. When we then account for the phraseologies occurring with these words, it could be expected that well over half of the text is covered.

I tested this supposition by taking a single randomly-chosen essay text (PIR3_008_R from POL/IR-X) and in it highlighting all phraseologies which had been semantically categorised as part of my analysis (phraseologies categorised as 'other' were not highlighted). The PIR3_008_R essay text, marked up as described in more detail below, can be found in Appendix 3. The first two paragraphs are also reproduced in Figure 4.1 below. Phraseologies for *of*, *that*, *as*, and *this* have been highlighted in different colours, and phraseologies for *and* have been underlined. Where there are overlaps in phraseologies, or one phraseology embedded within another, this has been indicated as far as possible by 'split' highlighting of words in two colours. Where one phraseology for *and* is embedded within another phraseology for *and* (N.B. there are no examples of this in Figure 4 but there are in the whole text in Appendix 3), double underlining is used to show this (N.B. There are also a very small number of words and phrases which have been highlighted in the essay in Appendix 3 by use of bold red italicised font that are not part of the current

phraseological study – these have been highlighted to demonstrate the comparatively low frequency in the text of items that were the foci of other recent examinations of student writing (Thompson, 2009; Nesi and Gardner, 2012; Durant, 2015) - these will be more fully explained and discussed in section 5.4.1 of Chapter 5).

Foucauldian accounts resistance [sic] stem from a re-conceptualization of power in opposition to ideas that power is located solely in the state apparatus, is the method of subjugating a population is the possession of a government or generally acts from rulers to the ruled in a downward action. Instead it is seen as being something existing everywhere, originating from innumerable points and moving in a network between numerous subjects. Resistance is seen as being not exterior to but within the power network itself, resistance is not against power but inherent in relationships of power. Using Foucault's re-conceptualization of power and resistance I shall argue against Marxist and Anarchist ideologies of resistance that characterise power as being the possession of a state and political resistance as being either the seizure or abolition of power centred within a state. I shall make the case that forms of resistance should learn from Foucauldian notions of power and resistance, recognise the true nature of power as inescapable and embrace forms of resistance that are heterogeneous in origins and local in target, directed towards interstices of dominating power relations.

Firstly let us examine the traditional theories of Marxism and Anarchism, how they conceive of the state, revolutionary activity and how they see the role of resistance, beginning with Marxism. For Marxism power is located solely within the state and in the hands of the bourgeoisie. As Foucault puts it, Marxism holds a juridico-economic conception of power as 'a right which can be possessed in the way one possesses a commodity' (Foucault. 2003: 13). The bourgeoisie, the 'class of modern Capitalists, owners of the means of social production and employers of wage labour' (Marx. 1967: 79) are the sole possessors of power.

Figure 4: The first two paragraphs of PIR3_008_R marked up for of, and, that, as & this phraseologies.

As can be seen in Figure 4.1, the result of this process is visually very striking. Only three sentences in this 2,693-word essay contained none of the phraseologies (and, if not for a small essay-writer error, the omission of the word *of* at the beginning of the first sentence of the introduction, there would be yet another *of* phraseology to highlight). Most sentences contained more than one. A count of words in the first paragraph, which is reasonably representative of the amount of phraseology

highlighted in other paragraphs in the text, finds that 146 of the 189 words are part of identified phraseologies, meaning coverage of more than 77% of the paragraph. It can thus be argued that from a pedagogical perspective, with regard to the teaching of academic writing in the disciplines and genre that are the focus of this project, the focus on *of*, *and*, *that*, *as*, and *this* is extremely fruitful.

4.2 Analysis

As outlined in section 3.7., the starting point with each target keyword was an analysis of a 300-line random sample taken from a corpus created by pooling all four 'disciplinary' and institutional sub-corpora. Once firm categories had been established, an analysis of a 300-line random sample from each 'disciplinary' and institutional sub-corpus was undertaken applying these categories. If further as yet unidentified categories emerge at this point, they were added.

Inherent in this process, because of its bottom-up qualitative nature, is the risk of ambiguities or overlaps in the categories formulated. Where they occur these issues have been acknowledged and discussed. There are also qualitative decisions to be made for each target keyword regarding the level of delicacy attempted in terms of divisions and consequent categories within the sample data. For proportionately larger groupings within a sample, deeper investigation and further sub-categorisation may be warranted, but this may be less useful or indeed practically possible for smaller groupings. Again, the decisions made regarding level of delicacy in categorisation are discussed as part of the analysis within this chapter. Table 22 below provides a reference to the reader for the labelling conventions used within the Chapter for semantic motifs, semantic sequences and concordance lines.

| Labelling convention | Meaning |
|---|--|
| 'single speech marks' | Semantic motif |
| <i>bold italics</i> | <i>Grammatical keyword</i> within a semantic sequence |
| BOLD FACE SMALL CAPS | SEMANTIC ELEMENT within a semantic sequence |
| + | Showing the <i>fixed-order relationship</i> between SEMANTIC ELEMENTS in a particular semantic sequence |
| & | Showing a <i>flexible order</i> for SEMANTIC ELEMENTS within a particular semantic sequence |
| (SEMANTIC ELEMENT IN BRACKETS) | an <i>optional SEMANTIC ELEMENT</i> |
| <u>underlined word or phrase</u> | Realisation of one semantic element in a concordance line. [N.B. A break in underlining distinguishes different elements within the line. Also, rather than using three dots ('...') I have indicated division of a semantic sequence across a longer stretch of text by not underlining the piece of text which is not considered to be part of the semantic sequence.] |
| highlighted word | Used to further highlight any salient features of phraseologies where useful |

Table 22: Labelling conventions for semantic motifs and semantic sequences

In some cases, a semantic sequence may go beyond the parameters of the concordance line given by *WordSmith*. This sometimes leads to ambiguity with regards to how to categorise the line. In such cases I have located the text file identified as the source of this line and used the 'Find' function to obtain the relevant extra text. I found this method more efficient than expanding the line within the *WordSmith* programme because I transferred my samples to Excel for analysis and manipulation; my corpora are of a small enough size that the tracing of individual source texts is an easy and quick process. This is done for all lines used in this Chapter as examples of semantic sequences, but, in the Excel documents containing the full 300-line samples, done only in cases where it has been deemed necessary to eliminate ambiguity in terms of how lines should be categorised.

In the presentation of the results below, between three and five concordance line examples are provided for each semantic sequence described. The exact number is decided in terms of how many lines are deemed necessary to give the reader sufficient insight into the logic and range of a particular grouping. In all cases where a sequence was identified in the first pooled random sample, the examples are taken from this sample. When examples for a sequence are drawn from any of the four disciplinary/institutional sub-corpora, this will be clearly indicated.

The results of the concordance analysis of each target keyword are presented below. First, semantic categories determined from analysis of the pooled sample are presented. Following from this each individual semantic category is explained and illustrated with examples. Finally, the frequencies across all four disciplinary/institutional sub-corpora are presented and continuities and variations in frequency across sub corpora are identified.

4.3 Of

4.3.1 Semantic categories for *of*

In discussing *of* it becomes extremely difficult to completely avoid discussion of grammatical properties; all the phraseologies for this keyword that occur frequently enough to be named are realised through ‘noun/noun phrase + *of* + noun/noun phrase’ nominalisations. These make up just fewer than 85% of the pooled random sample and just under or just over 90% of the lines in all four of the disciplinary/institutional sub-corpora. There are no semantic motifs for *of*; instead, there is a complex, at times overlapping taxonomy of semantic sequences. Fourteen of these groupings appear at least four times in either the pooled sample or in one or more of the samples from disciplinary/institutional sub-corpora. They are listed in order of frequency in Table 23 below.

| Category | Number of instances in 300-line sample | Percentage of 300-line sample |
|---|--|-------------------------------|
| 1. PROPERTY + <i>of</i> + PHENOMENON* | 55 | 18.33% |
| 2. PROCESS + <i>of</i> + OBJECT* | 39 | 13.00% |
| 3. PROCESS + <i>of</i> + ACTOR* | 28 | 9.33% |
| 4. CONCEPTUALISATION + <i>of</i> + PHENOMENON* | 16 | 5.33% |
| 5. PROPORTION OR AMOUNT + <i>of</i> + PHENOMENON | 18 | 6.00% |
| 6. QUALITY + <i>of</i> + PHENOMENON* | 16 | 5.33% |
| 7. PHENOMENON + <i>of</i> + INSTANCE OR NATURE | 15 | 5.00% |
| 8. CONCEPTUAL LABELLING + <i>of</i> + CONCEPT | 12 | 4.00% |
| 9. MODE + <i>of</i> + PROCESS | 11 | 3.67% |
| 10. AUTHORITY + <i>of</i> + DOMAIN* | 8 | 2.67% |
| 11. PHENOMENON + <i>of</i> + TIME* | 6 | 2.00% |
| 12. RESULT + <i>of</i> + PHENOMENON | 5 | 1.67% |
| 13. TEXT + <i>of</i> + CONTENT* | 4 | 1.33% |
| 14. CLASSIFICATION LABELLING + <i>of</i> + PHENOMENON | 2 | 0.67% |
| 15. Fixed phrases (including names) | 6 | 2.00% |
| 16. Other | 48 | 16.00% |
| 17. Book Titles | 11 | 3.67% |

Table 23: Semantic analysis of *of* in pooled sample (*categories used by Groom (2007))

The decision was made to exclude from categorisation phrases using *of* which I considered to be ‘fixed’ in nature such as ‘sphere of influence’ or ‘school of thought’; I interpreted these as operating as a single noun rather than a more flexible semantic sequence and also excluded the fixed names of people, places, events and institutions such as ‘Ministry of Defence’ or ‘the Liberation Tigers of Tamil Eelam’. Lines with patterns which occurred very infrequently (i.e. fewer than six times and thus less than 2% of the time) were grouped under ‘Other’. Unfortunately book titles appeared in a few lines of the random sample, always from the BAWE text files. This has a slight effect on the overall proportions for each semantic category. How this impact is mitigated is described below in section 4.1.3.

As already mentioned in Chapter 3, in comparison to the other keywords I look at in my study, my analysis of *of* has been particularly strongly influenced by Groom’s (2007) analysis because it was with *of* that he outlined his methodological procedure in detail. Many of the semantic sequence groupings and labels used

below are exactly the same as Groom's and where this is the case his influence has been clearly signalled and cited (in Table 4.4 this is indicated by a '*').

Appendix 4 contains a full copy of the completed analysis of the pooled sample for *of* (N.B. due to the University of Warwick regulations regarding the length of Appendices, the full analysis of concordance samples for all five keywords cannot be provided within this thesis).

4.3.2 Semantic sequences using *of*

4.3.2.1 PROPERTY + *of* + PHENOMENON

The largest grouping, making up just fewer than 20% of the lines in the pooled 300-line sample, is one in which the noun or noun phrase preceding *of* describes a 'property, attribute or feature' (Groom, 2007, p. 84) of the phenomenon following *of* (see Table 24 below). A test for this category is whether '*A of B*' can be reformulated as '*B has A*' (cf. Groom, 2007, p. 84); we can say the *the Bourbon Monarchy has appendages* and that *each ethnic conflict has a contextual background*.

| |
|--|
| ... <u>the appendages</u> <i>of</i> <u>the Bourbon Monarchy</u> ... |
| ... <u>the contextual background</u> <i>of</i> <u>each ethnic conflict</u> ... |
| ... <u>the lexicon</u> <i>of</i> <u>the feminist struggle</u> ... |
| ... <u>the personality</u> <i>of</i> <u>Willy Brandt</u> ... |
| ... <u>the weapons</u> <i>of</i> <u>democracy</u> ... |

Table 24: PROPERTY + *of* + PHENOMENON (1)

The broad, abstract nature of this semantic sequence means that it takes in a wide range of phenomena including particular people, places and events (*Willy Brandt, China, the New York protests*), institutions and regimes (*the international criminal court, the early Nasser regime*), human groupings (*Aborigines, the great families,*

the least advantaged), abstract social or political entities and movements (*the national state, institutions, the feminist struggle*), processes and states (*female medical aid, relations between the two*), and abstract concepts (*democracy, wealth*). It might be argued that further semantic division could have been made according to phenomenon type, so we could have, say, ‘PROPERTY + **of** + INDIVIDUAL’, ‘PROPERTY + **of** + HUMAN GROUP’ ‘PROPERTY + **of** + COUNTRY’ and so on. However, this proved impractical as it would have resulted in unworkably small sub-categories.

I have included in this grouping lines in which the ‘property’ can be seen as a ‘component’ of the ‘whole’ of the particular ‘phenomenon’ (see Table 25 below). Although, there were more than four such lines and thus they could arguably make a separate ‘COMPONENT + **of** + WHOLE PHENOMENON’ category, I decided that the overlap in the meaning of ‘property’ and ‘component’ meant this was not warranted.

| |
|--|
| ... <u>the surface of water</u> ... |
| ... <u>the outskirts of the great city</u> ... |
| <u>The most important part of his legacy</u> ... |

Table 25: PROPERTY + **of** + PHENOMENON (2)

I have also included in this grouping four lines in which the ‘property’ is the word *nature*. Although somewhat awkward in tone, *the nature of female medical aid* can be paraphrased as *female medical aid has a (particular) nature*. Lines which include a judgement about the *nature* such as *the **contradictory** nature of Britain’s foreign policy* have been grouped differently as part of the QUALITY + **of** + PHENOMENON category (described later); the rationale is that the best paraphrase for this phrase would be *Britain’s foreign policy is contradictory* (see Table 26 below).

| |
|---|
| ✓... <u>the nature of female medical aid</u> ... |
| ✗... <u>the contradictory nature of Britain’s foreign policy</u> ... |

Table 26: PROPERTY + **of** + PHENOMENON (3)

4.3.2.2 PROCESS + *of* + OBJECT

In the second largest category making up 13% of the pooled corpus sample the phenomenon following *of* is affected by the ‘process’ preceding *of* (cf. Groom, 2007, p. 85) (see Table 27 below). Here Groom’s choice of the systemic functional linguistic label ‘Process’ (Groom, 2007, p. 85) is consciously adopted. Groom’s decision to use ‘Object’ as a broad term ‘for describing things affected by processes’ rather than systemic functional linguistics’ more ‘complex nomenclature ... (i.e. “Range”, “Goal” etc.)’ (ibid.) is also followed.

A test for this semantic sequence is whether it can be paraphrased as a passive voice independent clause (cf. Groom, 2007, p. 86); so, for example, *Ecclesiastical condemnation of their preaching activities* would become *their preaching activities were condemned by the Church*.

... Ecclesiastical condemnation of their preaching activities ...
 ... creation of institutions at a regional level. ...
 ... deportation of undesirables ...
 ... display of consumer objects. ...
 ... loss of sovereignty ...

Table 27: PROCESS + *of* + OBJECT (1)

Three lines in this grouping are less prototypical. Preceding *of* in each of these instances it can be argued that there is not just a ‘process’ but a combination of ‘actor’ and ‘process’: *supporter of* is used instead of *support of*, *a ... driving force of* instead of *the driving of* and *[t]he authors of* instead of *the writing of* (see Table 28 below).

... an active supporter of German involvement and enmeshment in the Western spheres of political power. ...
 ... Oppenheimer was a great driving force of the Manhattan Project ...
 ... The authors of “Winstanley: A Case for the Man as he Said He Was” ...

Table 28: PROCESS + *of* + OBJECT (2)

These lines have been included in this category for two reasons. Firstly, there are too few instances for a separate category to be created and they are strongly enough related to be included rather than consigned to 'Other'. Secondly, they can be converted into passive voice independent clauses in the same way as the other members of the grouping: *German involvement and enmeshment in Western spheres of political power was supported*; *the Manhattan Project was driven (to a great extent by Oppenheimer)*; and, "Winstanley: A Case for the Man as He Said He Was" *was authored*'.

4.3.2.3 PROCESS + *of* + ACTOR

In this grouping a process precedes *of* and the actor in that process follows *of* (cf. Groom, 2007, pp. 90-91). These phraseologies should be able to be paraphrased as an independent clause; so, *a complete breakdown of effective government* can be reformulated as *effective government completely broke down*, and *intrusion of foreign states* can be reformulated as *foreign states intruded* (see Table 29 below). In line with Groom's finding for this semantic grouping (2007, p. 90), in my sample the majority of processes are ergative (Halliday, 1994). Many broadly express a sense of either growth (*development, evolution, formation, intensification, onset, rallying, rise, surge*) or decline (*slowing down, breakdown, death, failure*).

... a complete breakdown **of** effective government. ...
 ... the death **of** their leader Simon de Montfort. ...
 ... the development **of** socialism ...
 ... intrusion **of** foreign states ...

Table 29: PROCESS + *of* + ACTOR (1)

There are five lines I have included here which are arguably less prototypical for the grouping. Firstly, two lines with *function* as the 'process' are included (see Table 30 below). Their reformulation as independent clauses is a little awkward-sounding; *a function of knowledge* becomes *knowledge functions* and the "semiotic" *function of*

goods becomes *goods function “semiotically”*. Nonetheless, these lines have been deemed to fit best in this semantic grouping.

| |
|---|
| <p>... <u>a function of knowledge</u> ...</p> <p>... <u>the “semiotic” function of goods.</u> ...</p> |
|---|

Table 30: PROCESS + *of* + ACTOR (2)

Secondly, *the “sovereignty” of the people* is included in this category because it can be reformulated as *the people reign* if *sovereignty* is interpreted as a ‘process’ (see Table 31 below). It could also arguably be categorised as **PROPERTY + *of* + PHENOMENON**, i.e. *the people have “sovereignty”*, if *sovereignty* were interpreted as a ‘property’. The decision was made, however, that the sense of process is stronger than that of property so it has been grouped here. This line is, again, a useful illustration of the overlaps and ambiguities appearing at ‘the edges’ of many of these semantic groupings, where decisions about categorisation become markedly more interpretive.

| |
|---|
| <p>... <u>the “sovereignty of the people”</u> ...</p> |
|---|

Table 31: PROCESS + *of* + ACTOR (3)

Finally, two lines which include the word *role* preceding ***of*** have been included (see Table 32 below). Again, here it could be argued that *role* is not a ‘process’ but a ‘property’ and these lines should be part of the **PROPERTY + *of* + PHENOMENON**. However, it has been judged here that a more natural paraphrase of *role* in each case would not be *has* *a role* but, in fact, *plays* *a role*, and, thus, *role* is better encapsulated as a ‘process’: *an international institution plays a role, environmental policy plays a central role*. Similarly, *approach* in a line from the RHUL-POL/IR random sample has been interpreted as a ‘process’ rather than a ‘property’ of *modern police forces*.

... the role of an international institution ...
 ... the central role of environmental policy ...
 ... the approach of modern police forces ... [RHUL-PIR]

Table 32: PROCESS + *of* + ACTOR (4)

4.3.2.4 PROPORTION OR AMOUNT + *of* + PHENOMENON

The test for membership of this category (illustrated in Figure Table 33 below) is whether, '[h]ow many of?' or '[h]ow much of?' is asked with regards to the element following *of*, the answer will contain or consist of the element preceding *of*. So, the question '*[h]ow much of the cost?*' can be answered '*70% of it*', and the question '*[h]ow many of the Annaliste tools?*' can be answered '*all of them*'. In some lines included in this category, the answer to the 'how many/how much?' question is not a measurable proportion or number, but still expresses a sense of relative amount. Thus, we could ask '*[h]ow much of society?*' or '*[h]ow many members of society?*' and receive the answer '*some members of it*'; likewise, '*[h]ow much relative power?*' could be answered '*a greater degree of it*' and '*[h]ow many wage earners?*' could be answered '*the general body of them*'.

... 70% of the cost ...
 ... all of the Annaliste tools ...
 ... some members of society ...
 ... a greater degree of relative power ...
 ... the general body of wage earners ...

Table 33: PROPORTION OR AMOUNT + *of* + PHENOMENON (1)

Included in this grouping is a small subset of lines in which the element preceding *of* contains *lack* or *absence* (see Table 34 below). Although perhaps not strictly expressing a proportion or amount, it was decided that these lines fitted best under this category; we can ask '*[h]ow much articulated military capability?*' Or '*[h]ow much solitude?*' and logically get the answer '*none*'.

... the lack of articulated military capabilities ...
 ... lack of solitude ...
 ... a lack of will ...

Table 34: PROPORTION OR AMOUNT + *of* + PHENOMENON (2)

The samples from the disciplinary/institutional sub-corpora included three lines containing *extent* (in Table 35 below) which were also deemed to best fit within this grouping in that they were broadly expressing a sense of proportion or amount.

... the extent of impracticality of such a claim ... [HIST-Y]
 ... the extent of this empowerment ... [POL/IR-Y]
 ... a certain extent of deterrent power ... [POL/IR-X]

Table 35: PROPORTION OR AMOUNT + *of* + PHENOMENON (3)

4.3.2.5 CONCEPTUALISATION + *of* + PHENOMENON

In this grouping (illustrated in Table 36 below) the element preceding *of* can be described as one of many possible conceptual framings for the phenomenon following *of*. So, for example, *the ideal of ergonomics* could logically be reformulated as *ergonomics is conceived of as an ideal*, but it could also be conceived of as a *principle* or a *scientific discipline*. Again, this semantic category exactly mirrors Groom's (cf. 2007, pp. 88-89).

... The scenario of 'all things being equal' ...
 ... the "prism of violence and terrorism" ...
 ... the ideal of ergonomics ...
 ... the case of Wiki Leaks ...
 ... the peace of normality ...

Table 36: CONCEPTUALISATION + *of* + PHENOMENON

4.3.2.6 QUALITY + *of* + PHENOMENON

In this semantic grouping the element preceding *of* is an evaluation or judgement of the element following *of* (cf. Groom, 2007, pp. 94-95) (see Table 37 below). Thus, if we ask the question ‘*what was the Albigensian crusade like?*’, we get the answer ‘*it was complex*’, or, if we ask the question ‘*how could ‘old luxury’ be evaluated?*’, we get the answer ‘*it was conspicuously wasteful and excessive*’.

| |
|---|
| ... <u>the complexities</u> <u>of</u> <u>the Albigensian crusade</u> ... |
| ... <u>the limitations</u> <u>of</u> <u>secular work</u> ... |
| ... <u>conspicuous waste and excesses</u> <u>of</u> <u>‘old luxury’</u> ... |
| ... <u>low level</u> <u>of</u> <u>technology</u> ... |

Table 37: QUALITY + *of* + PHENOMENON

4.3.2.7 PHENOMENON + *of* + PARTICULAR INSTANCE OR NATURE

In this semantic grouping (illustrated in Table 38 below) the element following *of* consists of information which provides further specificity and detail regarding the phenomenon preceding *of*. Thus, we get more information about exactly what type of *crowd* or *lifestyle*, or get more information about the exact nature of the *period* or *relations*.

| |
|---|
| ... <u>a crowd</u> <u>of</u> <u>enraged Congolese women</u> ... |
| ... <u>a lifestyle</u> <u>of</u> <u>elites</u> ... |
| ... <u>a period</u> <u>of</u> <u>violent “rocking”</u> ... |
| ... <u>relations</u> <u>of</u> <u>dependency</u> ... |

Table 38: PHENOMENON + *of* + PARTICULAR INSTANCE OR NATURE

A comparison of two lines which both contain in which the word *elites* follows *of* shows a key difference between this category and PROPERTY + *of* + PHENOMENON (see Table 39 below). In the first sequence *elites* describes the nature of the *lifestyle* preceding *of*. In the second sequence *the elites* are the holders or possessors of *the attitudes* preceding *of*.

| |
|--|
| PHENOMENON + <i>of</i> + PARTICULAR INSTANCE OR NATURE |
| ... <u>a lifestyle</u> of <u>elites</u> ... |
| PROPERTY + <i>of</i> + PHENOMENON |
| ... <u>the attitudes</u> of <u>the elites</u> ... |

Table 39: The difference between PHENOMENON + *of* + PARTICULAR INSTANCE OR NATURE and PROPERTY + *of* + PHENOMENON sequences

Similarly, two lines containing the near similes *example* and *case* usefully demonstrate the difference between this category and CONCEPTUALISATION + *of* + PHENOMENON (see Table 40 below). In the first phrase ‘*cross border relations* provides’ more detail about the ‘*example*’, i.e. it details *what exactly is being exemplified*. In the second, in contrast, ‘*Wiki Leaks*’ is being *conceptualised* as a ‘*case*’.

| |
|--|
| PHENOMENON + <i>of</i> + PARTICULAR INSTANCE OR NATURE |
| ... <u>an example</u> of <u>cross-border relations</u> ... |
| CONCEPTUALISATION + <i>of</i> + PHENOMENON |
| ... <u>the case</u> of <u>Wiki Leaks</u> ... |

Table 40: The difference between PHENOMENON + *of* + PARTICULAR INSTANCE OR NATURE and CONCEPTUALISATION + *of* + PHENOMENON sequences

4.3.2.8 CONCEPTUAL LABELLING + *of* + CONCEPT

In this grouping (illustrated in Table 41) the element preceding *of* is a word or phrase which functions as a simile for ‘conceptualisation’ and the element which follows *of* is the phenomenon framed as a ‘concept’.

... the concept of a just war itself ...
 ... the prominent mythology of martyrdom ...
 ... notions of identities ...
 ... Rawl's conception of "justice as fairness" ...
 ... a new understanding of power ...

Table 41: CONCEPTUAL LABELLING + *of* + CONCEPT

Included in this grouping are lines which could arguably be part of the **PROCESS + *of* + OBJECT** grouping: *Rawl's conception of "justice as fairness"* could be reformulated as *"justice as fairness" was conceived by Rawl*, and *a new understanding of power* could be reformulated as *power is newly understood*. However, *conception* and *understanding* were judged to be more strongly conceptual labels than processes, so the decision was made that they best fit under **CONCEPTUAL LABELLING + *of* + CONCEPT**. Again, these lines are a useful illustration of the overlaps at the less prototypical edges of some semantic groupings.

A useful further clarification of this grouping can be made by comparing two sequences containing the word *perception* preceding *of* which appear in two of the disciplinary/institutional sub-corpora samples and have been categorised differently (see Table 42). In the first sequence the element following *of* is a 'concept' so *perception* is best interpreted as a conceptual label. In the second sequence *Mexico* is the object of the process of *perception*.

| CONCEPTUAL LABELLING + <i>of</i> + CONCEPT |
|--|
| ... <u>the perception</u> of <u>limitless American power into international affairs</u> ... [HIST-X] |
| PROCESS + <i>of</i> + OBJECT |
| ... <u>the perception</u> of <u>Mexico</u> ... |

Table 42: The word *perception* in **CONCEPTUAL LABELLING + *of* + CONCEPT** and **PROCESS + *of* + OBJECT** sequences.

Similarly, lines including *sense* when this word is judged to mean 'idea' fall within the **CONCEPTUAL LABELLING + *of* + CONCEPT** grouping whereas lines in which *sense* is

judged to mean ‘feeling’ are more appropriately grouped under **PHENOMENON + *of* + PARTICULAR INSTANCE OR NATURE** (see Table 43 below).

| |
|--|
| CONCEPTUAL LABELLING + <i>of</i> + CONCEPT |
| ... <u>his readers’ sense of justice</u> ... [POL/IR-Y] ... <u>a realistic sense of our own fallible nature”</u> ... [POL/IR-Y] |
| PHENOMENON + <i>of</i> + PARTICULAR INSTANCE OR NATURE |
| ... <u>a sense of familiarity to the point of intimacy</u> ... [HIST-Y] ... <u>a sense of disorientation</u> ... [HIST-Y] |

Table 43: The word sense in **CONCEPTUAL LABELLING + *of* + CONCEPT** and **PHENOMENON + *of* + PARTICULAR INSTANCE OR NATURE** sequences.

4.3.2.9 **MODE + *of* + PROCESS**

This category might if it were smaller be subsumed within **PHENOMENON + *of* + PARTICULAR INSTANCE OR NATURE**. The element preceding ***of*** contains a simile of ‘mode’ whilst the element following ***of*** represents a process (see Table 44 below).

| |
|--|
| ... <u>every form of punishment</u> <u>the style of painting known as Neoclassicism</u> <u>non-violent means of mediation and integration</u> <u>a system of inclusion and exclusion</u> <u>a method of research</u> ... |
|--|

Table 44: **MODE LABEL + *of* + PROCESS**

Here again, further useful clarification of this semantic grouping is provided by a comparison of two sequences which share a word, *system*, but are categorised differently (see Table 45 below). In the first sequence the word *system* denotes one ‘mode’ among many for completing the ‘process’ of *identifying conflict as genocide*. However, in the second sequence the element following ***of***, *inclusion and exclusion*, is not a ‘process’ for which there are various choices of ‘mode’ but is instead providing specific detail about the nature of the *system* preceding ***of***.

| |
|--|
| MODE + <i>of</i> + PROCESS |
| ... a deficient system of <u>identifying conflict as genocide</u> ... |
| PHENOMENON + <i>of</i> + PARTICULAR INSTANCE OR NATURE |
| ... a system of <u>inclusion and exclusion</u> ... |

Table 45: The word system in MODE + *of* + PROCESS and PHENOMENON + *of* + PARTICULAR INSTANCE OR NATURE sequences.

4.3.2.10 AUTHORITY + *of* + DOMAIN

This is a grouping found in Groom's study (2007, p. 87) in which the element preceding *of* describes 'a person or group holding a position of social power, responsibility or esteem', and the element following *of* describes 'the domain over which this individual or group has authority' (ibid.) (illustrated in Table 46 below).

| |
|--|
| ... <u>the father</u> of <u>economics</u> ... |
| ... <u>the previous Byzantine rulers</u> of <u>Semitic lands</u> |
| ... <u>the head</u> of <u>the household</u> |

Table 46: AUTHORITY + *of* + DOMAIN

4.3.2.11 PHENOMENON + *of* + TIME

Again, this grouping (illustrated in Table 47), the elements of which need no further explanation than that given in the label, appears in Groom's study (2007, p. 93).

| |
|--|
| ... <u>The Fourth Lateran Council</u> of <u>1216</u> ... |
| ... <u>Iran-Iraq war</u> of <u>1980-88</u> |

Table 47: PHENOMENON + *of* + TIME

4.3.2.12 RESULT + *of* + PHENOMENON

This is a grouping which, if at all smaller, could be subsumed within **PROPERTY + *of* + PHENOMENON**. Again, the elements need no further explanation than that given by the label (see Table 48 below).

... The effect of this ...
... the limited impact of his socialist policies ...
... the influence of the railway ...
... such deep influences of religious conviction ...

Table 48: RESULT + *of* + PHENOMENON

4.3.2.13 TEXT + *of* + CONTENT

In this semantic category which is present in Groom's study (2007, pp. 86-87) the element preceding *of* 'describes a text or text type (or a series of texts or text types)' while the element following *of* 'indicates the content of the text(s) or text type(s) in question' (p. 86) (see Table 49 below). I have included *David's sketch of Marie Antoinette* here stretching the definition of 'text' to also take in visual representations of content.

... Feminist accounts of women's agency in conflicts ...
... the history of the interaction between Mastery and Slavery ...
... a summary of Galenic medicine practice ...
... David's sketch of Marie Antoinette ...

Table 49: TEXT + *of* + CONTENT

4.3.2.14 CLASSIFICATION LABELLING + *of* + PHENOMENON

This is a tiny grouping which failed to fit under any other semantic category described. There were only two instances of this in the pooled sample but six instances were found in one of the disciplinary/institutional samples. Classifying nouns in singular or plural form precede *of*: *type, kind, category, class* and *strand* (see Table 50 below).

| |
|--|
| ... <u>many other types of knowledge</u> ... |
| ... <u>two types of worker</u> ... |
| ... <u>more modern economic strands of thought</u> ... [from HIST-X] |

Table 50: CLASSIFICATION LABELLING + *of* + PHENOMENON

A line containing the word *fields* has been placed within this semantic grouping with the plural form of *field* in the context of this particular sequence being interpreted as a classifying label. However, two other sequences which also contain *field* were categorised as **CONCEPTUALISATION + *of* + PHENOMENON** sequences; here *field* has been interpreted as one way of conceiving of the elements following *of*, *power relations* and *medicine* (see Table 51 below).

| |
|--|
| CLASSIFICATION LABELLING + <i>of</i> + PHENOMENON |
| ... <u>fields of knowledge</u> ... (from HIST-Y) |
| CONCEPTUALISATION + <i>of</i> + PHENOMENON |
| ... <u>the strategic field of power relations</u> ... (POL/IR-X) |
| ... <u>the field of medicine</u> ... (HIST-X) |

Table 51: The word field(s) in **CLASSIFICATION LABELLING + *of* + PHENOMENON** and **CONCEPTUALISATION + *of* + PHENOMENON** sequences.

4.3.3 Comparison/contrast across disciplinary and institutional sub-corpora

The issue of book and article titles appearing in the two BAWE-sourced disciplinary/institutional sub-corpora, HIST-Y and POL/IR-Y, had to be solved before a meaningful comparison of frequencies of semantic groupings for *of* across discipline and institution could be made. Rather than modifying the text files in the BAWE-sourced sub-corpora to eliminate book and article titles from concordance analyses and starting the sampling again, a process which would be extremely time consuming, it was decided to take the more time-efficient and arguably just as effective route of recalculating percentages after eliminating the book and article

titles from the total. Table 52 below shows the frequencies and percentages across the four disciplinary/institutional sub-corpora before this was done.

| Category | Pooled | | HIST -X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|---|--------|--------|---------|--------|--------|--------|----------|--------|----------|--------|
| 1. PROPERTY + <i>of</i> + PHENOMENON* | 55 | 18.33% | 47 | 15.67% | 36 | 12.00% | 37 | 12.33% | 41 | 13.67% |
| 2. PROCESS + <i>of</i> + OBJECT* | 39 | 13.00% | 61 | 20.33% | 56 | 18.67% | 57 | 19.00% | 33 | 11.00% |
| 3. PROCESS + <i>of</i> + ACTOR* | 28 | 9.33% | 25 | 8.33% | 22 | 7.33% | 30 | 10.00% | 22 | 7.33% |
| 4. PROPORTION or AMOUNT + <i>of</i> + PHENOMENON | 18 | 6.00% | 25 | 8.33% | 15 | 5.00% | 17 | 5.67% | 31 | 10.33% |
| 5. CONCEPTUALISATION + <i>of</i> + PHENOMENON* | 16 | 5.33% | 10 | 3.33% | 19 | 6.33% | 15 | 5.00% | 12 | 4.00% |
| 6. QUALITY + <i>of</i> + PHENOMENON* | 16 | 5.33% | 26 | 8.67% | 21 | 7.00% | 9 | 3.00% | 26 | 8.67% |
| 7. PHENOMENON + <i>of</i> + INSTANCE or NATURE | 15 | 5.00% | 26 | 8.67% | 27 | 9.00% | 48 | 16.00% | 28 | 9.33% |
| 8. CONCEPTUAL LABELLING + <i>of</i> + CONCEPT | 12 | 4.00% | 10 | 3.33% | 7 | 2.33% | 12 | 4.00% | 22 | 7.33% |
| 9. MODE + <i>of</i> + PROCESS | 12 | 4.00% | 4 | 1.33% | 5 | 1.67% | 3 | 1.00% | 1 | 0.33% |
| 10. AUTHORITY + <i>of</i> + DOMAIN* | 8 | 2.67% | 1 | 0.33% | 6 | 2.00% | 7 | 2.33% | 0 | 0% |
| 11. PHENOMENON + <i>of</i> + TIME* | 11 | 3.67% | 2 | 0.67% | 4 | 1.33% | 0 | 0% | 0 | 0% |
| 12. RESULT + <i>of</i> + PHENOMENON | 5 | 1.67% | 7 | 2.33% | 1 | 0.33% | 8 | 2.67% | 3 | 1.00% |
| 13. TEXT + <i>of</i> + CONTENT* | 4 | 1.33% | 8 | 2.67% | 3 | 1.00% | 3 | 1.00% | 0 | 0% |
| 14. CLASSIFICATION LABELLING + <i>of</i> + PHENOMENON | 2 | 0.67% | 5 | 1.67% | 3 | 1.00% | 6 | 2.00% | 4 | 1.33% |
| 15. Fixed phrases (including names) | 6 | 2.00% | 11 | 3.67% | 6 | 2.00% | 9 | 3.00% | 7 | 2.33% |
| 16. Other | 48 | 16.00% | 32 | 10.67% | 40 | 13.33% | 38 | 12.67% | 49 | 16.33% |
| 17. Book Titles | 11 | 3.67% | 0 | 0% | 29 | 9.67% | 1 | 0.33% | 21 | 7.00% |

Table 52: Semantic sequences for *of* - frequency and percentage in the pooled and disciplinary/institutional sub-corpora samples

The total number of lines was reduced to 271 in HIST-Y after subtracting 29 lines containing book titles, and 266 in POL/IR-Y after subtracting 34 book titles. New percentages were then calculated and are shown in Table 53 below. These percentage adjustments involved turned out to be very slight and to have little impact on the differences which were revealed across four sub-corpora.

| Category | Pooled | | HIST-X | | HIST-Y (adjusted - % of 271) | | POL/IR-X | | POL/IR-Y (adjusted - % of 279) | |
|---|--------|--------|--------|--------|---------------------------------|--------|----------|--------|-----------------------------------|--------|
| 1. PROPERTY + <i>of</i> + PHENOMENON* | 55 | 18.33% | 47 | 15.67% | 36 | 13.28% | 37 | 12.33% | 41 | 14.70% |
| 2. PROCESS + <i>of</i> + OBJECT* | 39 | 13.00% | 61 | 20.33% | 56 | 20.66% | 57 | 19.00% | 33 | 11.83% |
| 3. PROCESS + <i>of</i> + ACTOR* | 28 | 9.33% | 25 | 8.33% | 22 | 8.12% | 30 | 10.00% | 22 | 7.89% |
| 4. PROPORTION or AMOUNT + <i>of</i> + PHENOMENON | 18 | 6.00% | 25 | 8.33% | 15 | 5.54% | 17 | 5.67% | 31 | 11.11% |
| 5. CONCEPTUALISATION + <i>of</i> + PHENOMENON* | 16 | 5.33% | 10 | 3.33% | 19 | 7.01% | 15 | 5.00% | 12 | 4.30% |
| 6. QUALITY + <i>of</i> + PHENOMENON* | 16 | 5.33% | 26 | 8.67% | 21 | 7.75% | 9 | 3.00% | 26 | 9.32% |
| 7. PHENOMENON + <i>of</i> + INSTANCE or NATURE | 15 | 5.00% | 26 | 8.67% | 27 | 9.96% | 48 | 16.00% | 28 | 10.04% |
| 8. CONCEPTUAL LABELLING + <i>of</i> + CONCEPT | 12 | 4.00% | 10 | 3.33% | 7 | 2.58% | 12 | 4.00% | 22 | 7.89% |
| 9. MODE + <i>of</i> + PROCESS | 11 | 3.67% | 4 | 1.33% | 5 | 1.85% | 3 | 1.00% | 1 | 0.36% |
| 10. AUTHORITY + <i>of</i> + DOMAIN* | 8 | 2.67% | 1 | 0.33% | 6 | 2.21% | 7 | 2.33% | 0 | 0% |
| 11. PHENOMENON + <i>of</i> + TIME* | 6 | 2.00% | 2 | 0.67% | 4 | 1.48% | 0 | 0% | 0 | 0% |
| 12. RESULT + <i>of</i> + PHENOMENON | 5 | 1.67% | 7 | 2.33% | 1 | 0.37% | 8 | 2.67% | 3 | 1.08% |
| 13. TEXT + <i>of</i> + CONTENT* | 4 | 1.33% | 8 | 2.67% | 3 | 1.10% | 3 | 1.00% | 0 | 0% |
| 14. CLASSIFICATION LABELLING + <i>of</i> + PHENOMENON | 2 | 0.67% | 5 | 1.67% | 3 | 1.10% | 6 | 2.00% | 4 | 1.43% |
| 15. Fixed phrases (including names) | 6 | 2.00% | 11 | 3.67% | 6 | 2.21% | 9 | 3.00% | 7 | 2.51% |
| 16. Other | 48 | 16.00% | 32 | 10.67% | 40 | 14.76% | 38 | 12.67% | 49 | 17.56% |
| 17. Book Titles | 11 | 3.67% | 0 | 0% | 29 | 9.67% | 1 | 0.33% | 24 | 7.00% |

Table 53: Semantic sequences for *of* - frequency and percentage in the pooled and disciplinary/institutional sub-corpora samples – adjusted percentages

4.3.3.1 Similarities

As this is the first keyword to be analysed, it is worth here briefly explaining what is meant by ‘similarity’ within the context of this study. The study uses the notion of ‘similarity’ in a relatively generous sense to describe all instances in which a semantic pattern occurs in reasonable numbers, at least 2%, across all sub-corpora. All such cases will be identified and when similarities are particularly strong, with relative proportions of particular semantic patterns being very close across sub-corpora, this will also be highlighted.

With regards to **of** the overall picture is one of similarities rather than differences across discipline and institution. The four sub-corpora exhibit similar patterns of frequency across the fourteen semantic sequences ranging from larger proportions at the top of the table to much smaller proportions from sequence 9 (**MODE + of + PROCESS**) onwards. Frequencies and percentages are particularly similar for two of the largest groupings, **PROPERTY + of + PHENOMENON** and **PROCESS + of + ACTOR** sequences.

Two potential differences were identified for two of the smaller categories. There was a potential disciplinary difference across the first samples for **PHENOMENON + of + TIME**, with no occurrences in the POL/IR sub-corpora but a few each in both of the History sub-corpora. However, this difference disappeared when instances were counted in second random samples as can be seen in Table 54 below.

| PHENOMENON + of + TIME | HIST - rhul | | HIST – BAWE (adjusted - % of 287) | | POL/IR - rhul | | POL/IR – BAWE (adjusted - % of 281) | |
|-------------------------------|--------------------|-------|--|-------|----------------------|-------|--|-------|
| Sample 1 | 2 | 0.67% | 4 | 1.48% | 0 | 0% | 0 | 0% |
| Sample 2 | 3 | 1.00% | 1 | 0.35% | 2 | 0.67% | 2 | 0.71% |

Table 54: Second 300-line disciplinary/institutional samples for **PHENOMENON + of + TIME**

Similarly, a potential institutional difference, greater incidence of **RESULT + of + PHENOMENON** lines in the two Institution X sub-corpora, disappeared in the second sample (see Table 55 below).

| RESULT + <i>of</i> + PHENOMENON | HIST - rhul | | HIST – BAWE (adjusted - % of 287) | | POL/IR - rhul | | POL/IR – BAWE (adjusted - % of 281) | |
|---------------------------------|-------------|-------|---|-------|---------------|-------|---|-------|
| Sample 1 | 7 | 2.33% | 1 | 0.37% | 8 | 2.67% | 3 | 1.08% |
| Sample 2 | 0 | 0% | 4 | 1.39% | 4 | 1.33% | 4 | 1.42% |

Table 55: Second 300-line disciplinary/institutional samples for RESULT + *of* + PHENOMENON

4.3.3.2 Variations

The only noticeable differences to be found are cases where frequencies are larger or smaller in one particular disciplinary/institutional sub-corpora. Instances of **PROCESS + *of* + OBJECT** phraseologies in the POL/IR-Y sub-corpora at 11.83% of the total are almost half that in the other three sub-corpora, whilst instances of **CONCEPTUAL LABELLING + *of* + CONCEPT** phraseologies are double that in the other sub-corpora. There are a noticeably larger proportion of **CONCEPTUALISATION + *of* + PHENOMENON** phraseologies in HIST-Y and of **PHENOMENON + *of* + PARTICULAR INSTANCE OR NATURE** phraseologies in the POL/IR-X.

4.4 And

4.4.1 Semantic categories for *and*

Sematic groupings for *and* in the pooled random sample are listed in Table 56 below. A little more than three quarters of the phraseologies for *and* in the pooled sample can be categorised under the umbrella semantic motif of ‘listed phenomena’. Three other semantic motifs emerged in analysis of the pooled sample: ‘addition’, making up approximately 10% of the pooled sample, and ‘consequence’ and phraseologies with *between* each making up between 3% and 4% of the pooled sample.

| Category | Number of instances in 300-line sample | Percentage of 300-line sample |
|---|--|-------------------------------|
| 'listed phenomena' | | |
| 1 PROCESS + <i>and</i> + PROCESS | 66 | 22.00% |
| 2 ABSTRACT CONCEPT + <i>and</i> + ABSTRACT CONCEPT | 46 | 15.33% |
| 3 QUALITY + <i>and</i> + QUALITY | 27 | 9.00% |
| 4 CLASSIFIER + <i>and</i> + CLASSIFIER | 19 | 6.33% |
| 5 'listed people' | 19 | 6.33% |
| 6 LOCATION + <i>and</i> + LOCATION | 14 | 4.67% |
| 7 ARGUMENT + <i>and</i> + ARGUMENT | 6 | 2.00% |
| 8 PHENOMENON/PHENOMENA + <i>and</i> + ITS/THEIR CHARACTERISTICS | 4 | 1.33% |
| 9 'listed miscellany' | 30 | 10.00% |
| TOTAL | 231 | 77.00% |
| 10 'addition' | 25 | 8.33% |
| 11 'consequence' | 11 | 3.67% |
| 12 phraseologies with <i>between</i> | 10 | 3.33% |
| 13 Other | 2 | 0.67% |
| 14 book titles | 21 | 7.00% |

Table 56: Semantic analysis of *and* in pooled sample

Lines with patterns which occurred very infrequently (i.e. fewer than six times and thus less than 2% of the time) were grouped under 'Other'.

As with the concordance lines for *of*, book titles appeared in some lines of the random sample. Again, this always involved BAWE text files and had a slight effect on the overall proportions for each semantic category in the HIST-Y and POL/IR-Y sub-corpora. How this impact is mitigated is described below in section 4.2.6.

4.4.2 'Listed phenomena'

Determining which category each identified 'listed phenomena' phraseology belonged to required taking the narrowest possible boundaries for the phraseologies. This is illustrated in Figure Table 57 below.

| |
|--|
| ... <u>domestic</u> and <u>foreign</u> policies ... |
| ... relevant <u>conditions</u> and <u>factors</u> ... |
| ... <u>private</u> fortune and <u>state</u> power ... |

Table 57: Phraseology boundaries for *and*

In the first phraseology in Figure 4.30, the elimination of the word *policies* means the phraseology belongs to the **CLASSIFIER + *and* + CLASSIFIER** category and does not overlap with the **ABSTRACT CONCEPT + *and* + ABSTRACT CONCEPT** category. The second and third examples in Figure 4.30, both belonging to the **ABSTRACT CONCEPT + *and* + ABSTRACT CONCEPT** grouping, illustrate when it is and when it is not possible to narrow the boundaries of ‘listed phenomena’ phraseologies.

4.4.2.1 **PROCESS + *and* + PROCESS**

The semantic sequence **PROCESS + *and* + PROCESS** is the largest ‘listed phenomena’ grouping making up just over a fifth of the pooled sample is ‘listed processes’ (illustrated in Table 58 below). The **PROCESS** element within these sequences can consist of a verb phrase, as in the first three examples in Table 58, or a noun or noun phrase, as in the last two examples, illustrative of the fact that within this analytical procedure ‘**PROCESS**’ is a semantic rather than grammatical category descriptor. The third example in Figure 4.31, *readily accepted and well-conceived*, can be seen as a more ‘borderline’ example of this category in that the second **PROCESS** element, *well-conceived*, could be classified both as a **PROCESS** element or as a **QUALITY** element (see 4.4.2.3 below), and has been classified as the former because it is listed after an unambiguously **PROCESS** element. This is a good illustration of the qualitative decisions which must be made within this analytical process.

| |
|--|
| ... <u>preparing women for discharge</u> , and <u>procuring suitable situations for them</u> ... |
| ... <u>the successful mobilisation of jihad</u> and <u>the fusion of the military and the religious</u> ... |
| ... <u>readily accepted</u> and <u>well-conceived</u> ... |
| ... <u>an increase in poverty</u> and <u>an increase in theft</u> ... |
| ... <u>violence</u> and <u>provocation</u> ... |

Table 58: **PROCESS + *and* + PROCESS**

4.4.2.2 ABSTRACT CONCEPT + *and* + ABSTRACT CONCEPT

The sequence ABSTRACT CONCEPT + *and* + ABSTRACT CONCEPT (illustrated in Table 59 below) is the second largest ‘listed phenomena’ grouping making up fifteen percent of the pooled sample. A broad definition of ‘abstract’ phenomena as ‘an idea, quality, or state rather than a concrete object’ (Oxford English Dictionary) has been taken here.

| |
|---|
| ... <u>Language</u> and <u>culture</u> ... |
| ... <u>policies</u> and <u>political behaviour</u> ... |
| ... <u>forms</u> and <u>stages</u> ... |
| ... <u>Marxism</u> and <u>Anarchism</u> ... |
| ... <u>environmental problems</u> and <u>world poverty</u> ... |

Table 59: ABSTRACT CONCEPT + *and* + ABSTRACT CONCEPT

4.4.2.3 QUALITY + *and* + QUALITY

The third largest ‘listed phenomena’ grouping, constituting nine percent of the pooled sample, is the semantic sequence QUALITY + *and* + QUALITY (illustrated in Table 60 below). The QUALITY element in these sequences consists of a noun or noun phrase or an adjective.

| |
|--|
| ... <u>economic plenty</u> and <u>material luxury</u> ... |
| ... <u>corruption</u> and <u>predatoriness</u> ... |
| ... <u>talents</u> and <u>ability</u> ... |
| ... <u>demonic</u> and <u>irreligious</u> ... |
| ... <u>talented</u> and <u>influential</u> ... |

Table 60: QUALITY + *and* + QUALITY

4.4.2.4 CLASSIFIER + *and* + CLASSIFIER

The CLASSIFIER + *and* + CLASSIFIER GROUPING (illustrated in Table 61) could arguably be subsumed within the QUALITY + *that* + QUALITY grouping above considering the fact that the phraseologies in this grouping also perform a *characterising* function. However, they were differentiated on the basis that QUALITY + *that* + QUALITY sequences have an essentially evaluative purpose whereas the purpose of CLASSIFIER + *and* + CLASSIFIER sequences is to indicate where, temporally, geographically or in terms of abstract ‘realm’, phenomena can be placed.

| |
|---|
| ... <u>eighteenth</u> <i>and</i> <u>late-nineteenth</u> ... |
| ... <u>personal</u> , <u>academic</u> <i>and</i> <u>political</u> ... |
| ... <u>domestic</u> <i>and</i> <u>international</u> ... |
| ... <u>Western</u> <i>and</i> <u>Eastern</u> ... |

Table 61: CLASSIFIER + *and* + CLASSIFIER

4.4.2.5 ‘Listed people’

The grouping ‘listed people’, as illustrated below in Table 62, involves a range of semantic sequences taking in in both particular individuals or groups as well as more abstract ‘roles’ such as *the self-conscious and the other* and *trend setter and leader*.

| |
|--|
| INDIVIDUAL + <i>and</i> + INDIVIDUAL |
| ... <u>Bloch</u> and <u>Fevre</u> ... |
| INDIVIDUAL + <i>and</i> + GROUP |
| ... <u>Mao</u> and <u>the politburo</u> ... |
| [GROUP, GROUP, etc.] GROUP + <i>and</i> + GROUP |
| ... <u>doctors</u> , <u>apothecaries</u> and <u>grocers</u> <u>the Punjabi</u> , <u>Pashtun</u> , <u>Sindhi</u> , <u>Sariaki</u> , <u>Muhajirs</u> , and <u>Balochi</u> ... |
| ROLE + <i>and</i> + ROLE |
| ... <u>the self-conscious</u> and <u>the other</u> <u>trend setter</u> and <u>leader</u> ... |

Table 62: 'Listed people'

4.4.2.6 LOCATION + *and* + LOCATION

The majority of phraseologies within the LOCATION + *and* + LOCATION grouping (illustrated in Table 63) in the pooled sample are listed countries, regions or cities. A small minority, illustrated by the last two examples in Figure 4.36, refer to locations at a narrower level than these broad geographical categories.

| |
|--|
| ... <u>Britain</u> and <u>France</u> ... |
| ... <u>Inga</u> and <u>Shaba</u> ... |
| ... <u>the homes of the aristocracy</u> and <u>the gentlemen's clubs of St James and Pall Mall</u> ... |
| ... <u>house, street</u> and <u>land</u> ... |

Table 63: LOCATION + *and* + LOCATION

4.4.2.7 ARGUMENT + *and* + ARGUMENT

The ARGUMENT + *and* + ARGUMENT category, illustrated in Table 64, involves phraseologies which list two attributed propositions. In most cases each ARGUMENT element begins with the word *that*.

| |
|---|
| ... they develop the theory <u>that “institutions are created simply in response to state interests, and that their character is structured by the prevailing distribution of capabilities.”</u> ... |
| ... MacPherson essentially argues <u>that the Levellers 'always intended... a franchise excluding servants and alms-takers; and that they saw no inconsistency between this exclusion and their assertion of the natural right of every man to a vote</u> ... |
| ... Mills stated rightly <u>that "he who violates that right in another forfeits it in himself" and if the murderer forfeits his right to life by violating the right of another to life, then violating the right of another to life, then society is justified in imposing the worst (morally permissible) type of punishments on the murderer.</u> ... |

Table 64: ARGUMENT + *and* + ARGUMENT

4.4.2.8 PHENOMENON/PHENOMENA + *and* + ITS/THEIR CHARACTERISTICS

This is a very small category, illustrated in Table 65, making up less than two percent of the pooled sample and less than two percent of three out of four sub-corpora. It is included because it constitutes exactly two percent of one disciplinary/institutional sub-corpus, POL/IR-X. The element following *and* usually contains a singular or plural possessive pronoun, most often *its* or *their*.

| |
|---|
| ... <u>the land</u> and <u>its</u> rhythm. ... |
| ... <u>other advances in technology</u> and <u>what they</u> may hold for the future of warfare ... |
| ... <u>the international tensions</u> and <u>their</u> minimum level of intensity ... |

Table 65: PHENOMENON/PHENOMENA + *and* + ITS/THEIR CHARACTERISTICS

4.4.2.9 'Listed miscellany'

Almost ten percent of phraseologies within the pooled sample were judged not to fully fit within any of the other 'listed phenomena' groupings and relegated to 'listed miscellany' (illustrated in Table 66). Within this category there are potentially more meaningful categories but instances of phrases that fit each of these are too few in number; for example, if there were enough lines with similar types of meaning to the phrase *the atomic bomb and the hydrogen bomb* (in Figure 4.39 below) – there is in fact only one other, *food and fuel* - a **MATERIAL OBJECT + and + MATERIAL OBJECT** category could be formed.

| |
|---|
| ... <u>the atomic bomb</u> and <u>the hydrogen bomb</u> ... |
| ... <u>billions of dollars</u> and <u>hundreds of thousands of lives</u> ... |
| ... <u>the Song of the Cathar Wars</u> and <u>the Chronicle of William of Puylaurens</u> ... |
| ... <u>a scientist gone mad</u> and <u>a threat to world order</u> ... |

Table 66: PHENOMENON/PHENOMENA + **and** + PHENOMENON/PHENOMENA

4.4.3 'Addition'

The largest of the semantic groupings which do not come under the 'listed phenomena' umbrella is that of 'addition'. Table 67 illustrates how 'addition' is different from **PROCESS + and + PROCESS** sequences. A '**PROCESS + and + PROCESS**' phraseology is always *part* of an idea being expressed, grammatically, part of a clause. In contrast, in 'addition' phraseologies the **STATEMENT** and **ADDITIONAL STATEMENT** elements both consist of complete ideas, grammatically, whole clauses.

| |
|---|
| <p align="center">‘Listed processes’- PROCESS + <i>and</i> + PROCESS</p> <p>... In Germany Ranke's true philosophical perception was <u>readily accepted</u> and <u>well conceived</u>. ...</p> <p>... deteriorating living standards of the labouring classes lead to <u>an increase in poverty</u> and <u>an increase in theft</u> ...</p> |
| <p align="center">‘Addition’- STATEMENT + <i>and</i> + ADDITIONAL STATEMENT</p> <p>... The vision that made the Holy City of such importance was a legacy bequeathed to him by Nur al-Din, and <u>this became a focussing objective for Muslim action and triumph</u>.</p> <p>... Thus Jacoba was in all but name an elite and competent physician, and <u>her dismissal of other female healers as ‘ignorant women and inexperienced fools’ confirms this</u>. ...</p> <p>... The victim's wounds are visible, and <u>the lower half of the Marat is somewhat dominated by the redness of blood</u>. ...</p> |

Table 67: Difference between ‘Listed processes’ and ‘Addition’ motifs for *and*

4.4.4 ‘Consequence’

In ‘consequence’ phraseologies, which constitute just under four percent of the pooled sample, the element preceding ***and*** represents a cause the main element following ***and*** is a consequence. In sequences realising this motif a linguistic signal of a cause-consequence relationship usually directly follows ***and*** preceding the **CONSEQUENCE** element. In the pooled sample this **SIGNAL OF CONSEQUENCE** element is realised four times by *therefore*, three times by *so*, twice by *consequently* and once each by *hence* and *thereby*. Sometimes, as illustrated by the final example in Table 68, the **SIGNAL OF CONSEQUENCE** element is embedded within the **CONSEQUENCE** element.

| |
|--|
| ... <u>he would attain perfect empathy with his historical agents and so understand the past in its own terms</u> ... |
| ... He seems to have seen his task as the ensuring the "eternalisation" of the oath, and <u>therefore</u> based the scene on a classical composition ... |
| ... <u>President Johnson in the planning of Operation Rolling Thunder with the Joint Chiefs of Staff acknowledged that the risk attached to the "China Factor" would never be eliminated and that it would therefore be appropriate to proceed with escalation on American terms in the wake of the Gulf of Tonkin crisis.</u> ... |

Table 68: CAUSE + *and* + SIGNAL OF CONSEQUENCE + CONSEQUENCE

4.4.5 Phraseologies with *between*

A grouping of ten lines making up just over three percent of the pooled sample consists of phraseologies containing *between* (illustrated in Table 69 below). These are not an entirely coherent semantic grouping, but are too small in number to warrant finer division. These phraseologies express 'scope', 'division' or 'a link' between the elements preceding and following *and*.

| |
|--|
| 'scope' |
| ... reform <i>between</i> <u>1850</u> and <u>1890</u> ... |
| ... the death of <i>between</i> <u>700</u> and <u>2,500</u> ... |
| 'division' |
| ... differentiation <i>between</i> <u>military institution</u> and <u>the person</u> ... |
| ... the Cold War <i>between</i> <u>the United States</u> and <u>the Soviet Union</u> ... |
| 'a link' |
| ... continuity ... <i>between</i> <u>it</u> and <u>previous media</u> ... |
| ... the bridge <i>between</i> <u>East</u> and <u>West</u> ... |

Table 69: Phraseologies with *between*

4.4.6 Comparison/contrast across disciplines and institutions

As with *of*, in order to allow a meaningful cross-sub-corpora comparison of frequencies for and semantic groupings for *and*, percentages had to be recalculated for both HIST-Y and POL/IR-Y after book titles has been eliminated. Table 70 below provides the raw numbers and percentages.

| | Pooled | | HIST-X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|--|--------|--------|--------|--------|--------|--------|----------|--------|----------|--------|
| ‘listed phenomena’ | | | | | | | | | | |
| 1 PROCESS + <i>and</i> + PROCESS | 66 | 22.00% | 70 | 23.33% | 66 | 22.00% | 74 | 24.67% | 71 | 23.67% |
| 2 ABSTRACT CONCEPT + <i>and</i> + ABSTRACT CONCEPT | 46 | 15.33% | 22 | 7.33% | 28 | 9.33% | 38 | 12.67% | 33 | 11.00% |
| 3 QUALITY + <i>and</i> + QUALITY | 27 | 9.00% | 36 | 12.00% | 23 | 7.67% | 30 | 10.00% | 23 | 7.67% |
| 4 CLASSIFIER + <i>and</i> + CLASSIFIER | 19 | 6.33% | 14 | 4.67% | 11 | 3.67% | 18 | 6.00% | 12 | 4.00% |
| 5 ‘listed people’ | 19 | 6.33% | 29 | 9.67% | 19 | 6.33% | 26 | 8.67% | 18 | 6.00% |
| 6 LOCATION + <i>and</i> + LOCATION | 14 | 4.67% | 21 | 7.00% | 5 | 1.67% | 8 | 2.67% | 9 | 3.00% |
| 7 ARGUMENT + <i>and</i> + ARGUMENT | 6 | 2.00% | 3 | 1.00% | 3 | 1.00% | 3 | 1.00% | 1 | 0.33% |
| 8 PHENOMENA + <i>and</i> + THEIR CHARACTERISTICS | 4 | 1.33% | 4 | 1.33% | 2 | 0.67% | 6 | 2.00% | 1 | 0.33% |
| 9 ‘listed miscellany’ | 30 | 10.00% | 36 | 12.00% | 45 | 15.00% | 57 | 19.00% | 45 | 15.00% |
| TOTAL ‘listed phenomena’ | 231 | 77.00% | 235 | 78.33% | 202 | 67.33% | 260 | 86.67% | 213 | 71.00% |
| 10 ‘addition’ | 25 | 8.33% | 53 | 17.67% | 30 | 10.00% | 23 | 7.67% | 32 | 10.67% |
| 11 ‘consequence’ | 11 | 3.67% | 7 | 2.33% | 6 | 2.00% | 5 | 1.67% | 6 | 2.00% |
| 12 phraseologies with <i>between</i> | 10 | 3.33% | 4 | 1.33% | 5 | 1.67% | 7 | 2.33% | 4 | 1.33% |
| 13 uncategorised | 2 | 0.67% | 1 | 0.33% | 4 | 1.33% | 4 | 1.33% | 4 | 1.33% |
| 14 book titles | 21 | 7.00% | | | 53 | 17.67% | 1 | 0.33% | 41 | 13.67% |

Table 70: Semantic sequences and motifs for *and* – frequency and percentage in pooled corpus and disciplinary/institutional sub-corpora

Table 71 shows adjusted percentages for HIST-Y and POL/IR-Y.

| | Pooled | | HIST-X | | HIST-Y [Percentages out of 247] | | POL/IR-X | | POL/IR-Y [Percentages out of 259] | |
|--|--------|--------|--------|--------|---------------------------------------|--------|----------|--------|---|--------|
| ‘listed phenomena’ | | | | | | | | | | |
| 1 PROCESS + <i>and</i> + PROCESS | 66 | 22.00% | 70 | 23.33% | 66 | 26.72% | 74 | 24.67% | 71 | 27.41% |
| 2 ABSTRACT CONCEPT + <i>and</i> + ABSTRACT CONCEPT | 46 | 15.33% | 22 | 7.33% | 28 | 11.34% | 38 | 12.67% | 33 | 12.74% |
| 3 QUALITY + <i>and</i> + QUALITY | 27 | 9.00% | 36 | 12.00% | 23 | 9.31% | 30 | 10.00% | 23 | 8.88% |
| 4 CLASSIFIER + <i>and</i> + CLASSIFIER | 19 | 6.33% | 14 | 4.67% | 11 | 4.45% | 18 | 6.00% | 12 | 4.63% |
| 5 ‘listed people’ | 19 | 6.33% | 29 | 9.67% | 19 | 7.69% | 26 | 8.67% | 18 | 6.95% |
| 6 LOCATION + <i>and</i> + LOCATION | 14 | 4.67% | 21 | 7.00% | 5 | 2.02% | 8 | 2.67% | 9 | 3.47% |
| 7 ARGUMENT + <i>and</i> + ARGUMENT | 6 | 2.00% | 3 | 1.00% | 3 | 1.21% | 3 | 1.00% | 1 | 0.40% |
| 8 PHENOMENA + <i>and</i> + THEIR CHARACTERISTICS | 4 | 1.33% | 4 | 1.33% | 2 | 0.81% | 6 | 2.00% | 1 | 0.40% |
| 9 ‘listed miscellany’ | 30 | 10.00% | 36 | 12.00% | 45 | 18.22% | 57 | 19.00% | 45 | 17.37% |
| TOTAL ‘listed phenomena’ | 231 | 77.00% | 235 | 78.33% | 202 | 81.78% | 260 | 86.67% | 213 | 82.24% |
| 10 ‘addition’ | 25 | 8.33% | 53 | 17.67% | 30 | 12.15% | 23 | 7.67% | 32 | 12.36% |
| 11 ‘consequence’ | 11 | 3.67% | 7 | 2.33% | 6 | 2.42% | 5 | 1.67% | 6 | 2.32% |
| 12 phraseologies with <i>between</i> | 10 | 3.33% | 4 | 1.33% | 5 | 2.02% | 7 | 2.33% | 4 | 1.54% |
| 13 uncategorised | 2 | 0.67% | 1 | 0.33% | 3 | 1.21% | 4 | 1.33% | 4 | 1.54% |
| 14 book titles | 21 | 7.00% | | | 53 | 17.67% | 1 | 0.33% | 41 | 13.67% |

Table 71: Semantic sequences and motifs for *and* – frequency and percentage in pooled corpus and disciplinary/institutional sub-corpora – adjusted percentages

4.4.6.1 Similarities

There are a large number of similarities in the semantic patterning of ***and*** across the disciplinary/institutional sub-corpora, the foremost being the dominance of 'listed phenomena' which make up between 78% and 87% of all phraseologies across the sub-corpora. Similar proportions of each sub-corpus are made up of the following sequences and motifs: PROCESS + ***and*** + PROCESS, ABSTRACT CONCEPT + ***and*** + ABSTRACT CONCEPT, QUALITY + ***and*** + QUALITY, CLASSIFIER + ***and*** + CLASSIFIER, and 'listed people'. The proportions of LOCATION + ***and*** + LOCATION and 'addition' phraseologies vary across

the four sub-corpora but nonetheless occur in noticeable numbers in each so in this way can be considered to represent further similarity.

4.4.6.2 Variations

In Table 4.10 there are four instances in which there is a very slightly higher frequency of a particular semantic grouping in the sub-corpora from both disciplines at 'X' institution. However, in each case, the difference is so slight as to not indicate a clear difference in emphasis: in the **PROCESS + *and* + PROCESS** category the difference is between 2% and 3% for already large grouping across all four sub-corpora and the picture is more of similarity than difference; for the **QUALITY + *and* + QUALITY**, 'listed people' and **PHENOMENA + *and* + THEIR CHARACTERISTICS** groupings the differences are even smaller. These very slight variations have therefore been deemed not worth further investigation by use of second random samples.

4.5 That

4.5.1 Semantic categories for *that*

More than half of the phraseologies for ***that*** in the pooled sample convey what has been termed 'voice' (see Table 72 below). For the purposes of this study 'voice' denotes 'what is being said about the field in focus'. In the discussion below, a distinction has been made between 'the voice of the essay writer' and 'the voice of others'.

'The voice of others' could arguably have been further subdivided according to whether the sequences conveyed the voice of *scholars of* the field, *actors in* the field (e.g. *President Bashir*) or *contemporary commentators on* the field (e.g. *the British press*). This, however, proved a lot more complicated than initially anticipated. It was not always completely clear which of these sub-categories a

particular sequence belonged to meaning that further subdivision of this semantic category would lead to grouping involving the extremely time-consuming process of not only expanding lines to examine more co-text but also checking reference lists for some of the essay text files. Furthermore, there was frequently overlap between who might be deemed a ‘scholar’ as opposed to a ‘field commentator’. This was further complicated by the question of whether, in an essay on the topic of, for example, Marx, the voice of Marx is ‘scholarly’ or ‘from the field’. Therefore, the decision was made to not sub-divide this grouping any further.

| Category | Number of instances in 300-line sample | Percentage of 300-line sample |
|---|--|-------------------------------|
| ‘Voice’ | | 55.00% |
| <u>The voice of others</u> | | |
| ‘Reporting the ideas of others’ | 107 | 35.67% |
| <u>The voice of the essay writer</u> | | |
| ‘Essay-writer argument’ | 33 | 11.00% |
| ‘Essay-writer interpretation’ | 25 | 8.33% |
| PHENOMENON + <i>that</i> + DETAIL | 70 | 23.33% |
| Fixed expression - <i>the fact that</i> | 13 | 4.33% |
| ‘Other’ | 52 | 17.33% |

Table 72: Semantic categories and motifs for *that*

Only two further clear semantic groupings were identified within the pooled sample for *that*, the semantic sequence ‘PHENOMENON + *that* + DETAIL’ and sequences involving the fixed expression *the fact that*.

4.5.2 ‘The voice of others’

‘The voice of others’ is realised through a range of semantic sequences. These sequences all share two common features: firstly, the element to the right of *that* is a ‘statement’ in the broad sense of ‘the voice of somebody expressing something’; secondly, the element or sequence of elements to the left of *that* imbue this statement with particular attributes, character and meaning within the essay text beyond that which can be deduced from the statement itself in isolation. In other

words, from the elements preceding **that** we understand how the statement following **that** is being framed within the essay. This framing can include some or all of the following: the author or authors of the statement, whether the statement is, among a wider range of possibilities, an argument, belief, report, suggestion or theory, how confidently or cautiously the statement is being made, and how the statement is evaluated by its audience (often the essay writer).

By far the largest ‘voice of others’ grouping is the ‘AUTHOR OF STATEMENT + CHARACTER OF STATEMENT + **that** + STATEMENT’ semantic sequence. The ‘AUTHOR’ element of this sequence is in most cases either a name (e.g. *Toulouse Tudela, President Johnson*) or a personal pronoun (*they, he, I*). The ‘CHARACTER OF STATEMENT’ element of this sequence is often realised through a reporting verb alone, but can consist of a longer phrase as illustrated in the third and fourth lines in Table 73 below; ‘*stated rightly*’ contains the essay-writer’s judgement of the statement, and ‘*went on to add*’ positions the statement as an addition to a previous statement.

| AUTHOR OF STATEMENT + CHARACTER OF STATEMENT + that + STATEMENT |
|---|
| ... <u>Mark freeman argues that historians tend to rate such social journalism as similar to imaginative literature of the time</u> ... |
| ... <u>Scholars of this theory believe that once cooperation amongst States is institutionalised, States would be reluctant to leave it,</u> ... |
| ... <u>Mills stated rightly that "he who violates that right in another forfeits it in himself"</u> ... |
| ... <u>He went on to add that "inclusion in the East Mark currency zone would have been the beginning of the end for Berlin".</u> ... |
| ... but as Willy Brandt put it in 1958, <u>"I maintain that the idea of local reunification would fail on the issue of currency, quite apart from anything else".</u> |
| ... |

Table 73: 'The voice of others' (1)

As can be seen from Table 74 below, a small number of phrases possessing slightly different grammatical structures than that of the vast majority have been nevertheless deemed members of this semantic sequence. In these cases the semantic sequence has been interpreted as still present but spanning a longer stretch of language.

| |
|--|
| <p>... <u>Banerjee</u> supports the arguments put forth by the Sunder Rajan and the advocate from Majlis by <u>claiming that Hindu identity provides both security but also female empowerment</u>: ...</p> <p>... <u>a significant proportion of American historians that incorrectly believe that their work follows in the footsteps of Rankle</u> ...</p> <p>... <u>Zagorin and Hill, who were of the opinion that Winstanley's pantheism, which is manifested in the "...identification of God with the immanent principle of reason pervading this world"</u> ...</p> <p>... <u>according to Capp "it is clear that something had crystalised from the flux, and that we are no longer dealing with isolated individuals"</u>. ...</p> <p>... <u>Rainsborough</u> whose response was far clearer, and arguably was 'the simplest claim for equality ever made in English history': 'The Levellers and Democracy', p. 175 and Thomas, 'The Levellers and the '...every man that is to live under a government ought first by his own consent to put himself under that government; and <u>I do think that the poorest man in England is not at all bound in a strict sense to that government that he hath not had a voice to put himself under...</u>' ...</p> |
|--|

Table 74: 'The voice of others' (2)

The second most frequent semantic sequence is 'CHARACTER OF STATEMENT + *that* + STATEMENT' which always occurs in the form of a nominalisation. A single element precedes *that*, which, again, provides one or more layers of framing for the statement following *that*. So, in the examples in Table 75 below *the wide range of arguments* tells us that the statement is one of a diverse plurality of arguments making the same point. In a small proportion of these sequences, six out of the total twenty six, the statement is explicitly attributed to an author, i.e. *his argument* and *Davis's hypothesis*. In most the author is absent.

| CHARACTER OF STATEMENT + <i>that</i> + STATEMENT |
|--|
| <p>... <u>the wide range of arguments that points to the necessity for the European Union to develop a viable, efficient and respected European Army</u>, ...</p> <p>... <u>the charge that the discourse of OWS and Foucault's ideas of resistance cannot be reconciled</u>. ...</p> <p>... <u>the Marxist assumption that 'to attempt to stage a socialist revolution before the bourgeois order had established itself would be an act of irresponsible utopianism, was bound to fail, and would thus delay the ultimate victory'</u>. ...</p> <p>... <u>his argument that the veil of ignorance appears to "exclude any attempt to maximise expectations which would lead toward utilitarianism"</u> ...</p> <p>... <u>Davis's hypothesis that the Ranters did not exist</u> ...</p> |

Table 75: 'The voice of others' (3)

A smaller grouping of seven sequences in total include either *is* or *was* preceding *that*. The structure of this semantic sequence often allows for a more complex

framing of the 'STATEMENT' to be made in within the 'CHARACTER OF STATEMENT' element such as can be seen in the first three sequences in Table 76 below.

| CHARACTER OF STATEMENT + <i>is/was</i> + that + STATEMENT |
|---|
| ... <u>The prevailing notion amongst the neo-liberal economists was that in future years these financial investments would prove invaluable for a developing economy.</u> ... |
| ... <u>the key part of this argument is not that the workers have the capacity to change the future, but that the present - the capitalist present - is considered by Veblen to be a "historical distortion."</u> ... |
| ... <u>the principal objection to this is that it is unfair to those who are working hard to earn a living for themselves</u> ... |
| <u>the suggestion is that globalisation is something above us</u> ... |

Table 76: 'The voice of others' (4)

Only five sequences follow the '*It* + CHARACTER OF STATEMENT + **that** + STATEMENT' pattern in Table 77 below. This sequence facilitates the framing of a statement without attributing it to a particular author. Its more common usage is in the context of essay-writer arguments as we shall see below.

| <i>It</i> + CHARACTER OF STATEMENT + that + STATEMENT |
|--|
| ... <u>It was also noted in the British press that the American army was over reliant on machines, and that its forces would be reluctant to dismount their vehicles</u> ... |
| ... <u>It is claimed that the Secretary General is the world's prime example of responsibility without power, which is not always understood.</u> ... |
| ... <u>it is generally agreed that Waldes was a rich merchant from Lyon</u> ... |
| <u>It has also been argued that this entails that all subsequent writings about the Orient</u> ... |
| ... <u>It is argued that the Tsar should have helped the situation by taking steps to de-politicise industrial relations</u> ... |

Table 77: 'The voice of others' (5)

4.5.3 'The voice of the essay writer'

In this semantic motif the statement following **that** is attributed to the essay writer. This motif can be divided into two clear sub-categories, 'arguing' and 'interpretation'.

4.5.3.1 'The voice of the essay writer – arguing'

In 'arguing' sequences the elements preceding and following **that** work together to create an opinion which can be clearly attributed to the essay writer. In the simplest of these the statement following **that** is explicitly labelled by the element preceding **that** as an argument in and of itself, e.g. '*[i]t will then be argued that ...*', '*I believe that ...*', an opinion, '*it is my opinion that ...*', and '*I think that ...*'.

In more complex instances the argument does not lie solely within the statement following **that**, but, instead, is created by the combination of both key elements so that the element preceding **that** has a more important function than mere labelling. In the sequence '[I]t is important to remember that there are structural as well as political factors', the argument can be paraphrased as '*the existence of both structural as well as political factors must be kept in mind*', and, in the sequence 'it would seem that Madame de Pompadour could never be popular' the argument can be paraphrased as '*Madame de Pompadour quite possibly could never be popular*'; in both cases the elements preceding **this**, paraphrased as '*must be kept in mind*' and '*possibly*', are a crucial part of the argument being made by the essay writer.

This motif is realised by three semantic sequences. The first two, exemplified in Table 78 and Table 80, are differentiated by the choice the essay writer has made to either impersonalise or personalise their argument.

The largest grouping involves semantic sequences which begin with the word ‘it’. This element serves to impersonalise the argument by avoiding the need for a personal pronoun.

| <i>It</i> + ESSAY-WRITER ANGLE ON STATEMENT + that + STATEMENT |
|--|
| ... <u>It is certain that at the heart of the development of ARPANET, was an outlook that saw the Net as participatory, democratic and ruled by none.</u> ... |
| ... <u>it is all the more crucial that the gesture which binds the individuals to its realisation is the over-arching motif of the painting.</u> ... |
| ... <u>it could be argued that Ranke came to this conclusion on the basis of earlier research into the f and u,</u> ... |
| ... <u>It would be gross oversimplification to say that his error caused the Berlin Wall to fall</u> ... |
| ... <u>It seems that African politics is not only endowed by its colonial heritage, but has rather developed into an amalgamation of different elements,</u> ... |

Table 78: ‘The voice of the essay writer - arguing’ (1)

These sequences are similar in appearance to the ‘*It* + CHARACTER OF STATEMENT + **that** + STATEMENT’ sequences that make up a small portion of ‘the voice of others’ grouping. However, there are usually obvious clues in the ‘CHARACTER OF STATEMENT’ and ‘ESSAY-WRITER ANGLE ON STATEMENT’ elements respectively which differentiate these two types of semantic sequence as either belonging to ‘the voice of others’ or ‘the voice of the essay writer’. This is illustrated in Table 79 below. In the ESSAY-WRITER ANGLE ON STATEMENT elements in the left-hand column either evaluative language, *clear*, *important*, *oversimplification* or hedging language, *would* and *seems* indicate essay-writer voice. In the right-hand column attribution, *the British press*, a negative connoting argument label, *claimed*, and the phrase *generally agreed* indicate a voice other than the essay writer.

| ESSAY-WRITER ANGLE ON STATEMENT element ('The voice of the essay writer') | CHARACTER OF STATEMENT element ('The voice of others') |
|---|--|
| <p>It <u>is clear</u> that ...</p> <p>... It <u>is important to note</u> that ...</p> <p>... it <u>could be argued</u> that that ...</p> <p>... It <u>would be gross oversimplification to say</u> that ...</p> <p>... It <u>seems</u> that ...</p> | <p>... It <u>was also noted in the British press</u> that the American army was over reliant on machines, and that ...</p> <p>... It <u>is claimed</u> that ...</p> <p>... it <u>is generally agreed</u> that ...</p> <p>... It <u>has also been argued</u> that ...</p> <p>... It <u>is argued</u> that ...</p> |

Table 79: ESSAY-WRITER ANGLE ON STATEMENT element compared with CHARACTER OF STATEMENT element

The last two CHARACTER OF STATEMENT elements, *[it] has also been argued* and *[i]t is argued* were more ambiguous; an examination of the co-text was required to determine that these lines fell in the 'voice of others' rather than 'voice of the essay writer' grouping.

The next semantic sequence (in Table 80 below) is found in only five of the thirty three instances of 'the voice of the essay writer – arguing'. In these sequences the personal pronouns 'I', referring to the essay writer alone, or 'we' a more inclusive reference to both the essay writer and reader, serve to personalise the argument. Again, the co-text needs to be examined in some instances to confirm that the 'I' refers to the essay-writer and is not part of a quote and thus belonging to 'the voice of others'; in two instances this was the case.

| I/we + ESSAY-WRITER ANGLE ON STATEMENT + <i>that</i> + STATEMENT |
|---|
| <p>... I <u>believe that the poor conditions the workers were faced with led them to need to air their grievances publicly</u> ...</p> <p>... I <u>realise that Scanlon is attempting to justify his contractualism,</u> ...</p> <p>... We <u>can observe that Private Manning is subject to both disciplinary and security procedures.</u> ...</p> <p>... We <u>must accept that we can never remove ourselves from it and that we can never do away with.</u> ...</p> |

Table 80: 'The voice of the essay writer - arguing' (2)

Two lines belong to a final semantic sequence (see Table 81) which contains an element referring to the essay text itself followed by the auxiliary verb *will*. It is clear (and confirmed by a quick examination of the co-text) these sequences appear near the beginning of the essay and serve to express all or part of the thesis of the essay.

| REFERENCE TO ESSAY TEXT ITSELF + ESSAY-WRITER ANGLE ON STATEMENT + <i>that</i> + STATEMENT |
|---|
| <p>... <u>this paper will argue that understandings and discussions of the East- focussing here particularly on India- continue to be influenced by the fundamental feature of Orientalism.</u> ...</p> <p>... <u>the essay will also illustrate that the Islamic threat is not a coherent threat and it faces more threat from within than it poses to the "West".</u> ...</p> |

Table 81: 'The voice of the essay writer - arguing' (3)

4.5.3.2 'The voice of the essay writer – interpretation'

This semantic grouping has been labelled 'interpretation' and differentiated from the 'argument' grouping because although the 'interpretation' involved is obviously also an argument of the essay writer, these sequences involve a different process. In these sequences (exemplified in Table 82 below) the phenomenon occurring to the left of *that* is presented as evidence for the interpretation to the right of *that*. A small range of verbs is used in the 'INTERPRETATIVE ANGLE' element which links the 'PHENOMENON AS EVIDENCE' with the 'INTERPRETATION'. In the 300-line pooled sample only six appeared: *demonstrate*, *illustrate*, *indicate*, *mean*, *show* and *suggest*.

| PHENOMENON AS EVIDENCE + INTERPRETIVE ANGLE + <i>that</i> + INTERPRETATION |
|---|
| <p>... the complexity of the relationship with Soviet Russia demonstrates <u>that the character of the revolution cannot be placed under one ideological term.</u> ...</p> <p>... This illustrates <u>that Bayarts thesis is very relevant in understanding the 'consumption' of the African State</u> ...</p> <p>... Germany's federal structure, complete with powerful state governments, mean[sic] <u>that 'the architecture of Germany's political system is far more complex than the structure of a centralized unitary state'</u> ...</p> <p>... evidence suggests <u>that parish priests, who would probably have been local and thus well versed in popular culture, often used them to provide comfort to their practitioners.</u> ...</p> |

Table 82: 'The voice of the essay writer – interpretation'

4.5.4 PHENOMENON + *that* + DETAIL

The semantic sequence, PHENOMENON + *that* + DETAIL exemplified in Table 83 below, makes up a large grouping, just over twenty three percent of lines in the pooled sample. The element preceding *that* is a phenomenon and the element following *this* provides more information or detail about this phenomenon. The phenomena are diverse in range including the relatively concrete *environmental groups, Paris and Marseille, other leaders* and *smoke* to the more abstract *ideas of human agency, equality, priority* and *themes*. The types of DETAIL in the element following *that* was even more diverse in range, so there was no workable way to subdivide this grouping any further.

... a rare insight into ancient historiography *that* would otherwise have been lost ...
... policy change *that* exceeds mere institutional mediation. ...
... Muslim figures *that* could act both as rulers and leaders ...
... the 'cult of the individual' *that* is still apparent in the present Western society ...
... the squalor *that* arose from London's poor drainage ...

Table 83: PHENOMENON + *that* + DETAIL

4.5.5 Fixed expression – *the fact that*

A final clear category within the pooled sample contains the fixed expression *the fact that* which occurs thirteen times. Included in the thirteen instances in this category are three in which the expression *the fact that* has been elided to *that* (as shown in Table 84 below) (N.B. ellipted instances were only included where *that* could very easily and naturally be expanded to *the fact* within its context).

... That David did not intend to compromise the Revolution's rejection of the Church is clear from the way in which ... it is with unadulterated eloquence of the body that Marat's sacrifice for the people is expressed ...

Table 84: Ellipsis of *the fact that* to 'That'

The broad function of this phrase is grammatical metaphor (Halliday, 1985) in the reformulation of a clausal structure as a nominalisation. However, two clear semantic patterns (exemplified in Table 85) can be identified within a subset of these lines. The first is a cause-effect phraseology in which ***the fact that*** is followed by a ‘cause’. Six lines follow this pattern in one instance of which ***the fact that*** occurs as part of the longer fixed expressions ***due to the fact that***. In two lines ***the fact that*** is followed by ‘evidence’ used in an interpretation.

| (EFFECT +) <i>the fact that</i> + CAUSE (+ EFFECT) |
|--|
| ... <u>the maximalists were never successful</u> , perhaps <i>due to the fact that</i> <u>they had a rather soft core and they "provided only revolutionary talk as a substitute for revolution.</u> ... |
| ... <i>the fact that</i> <u>the United States was accommodating the Franco regime as the Cold War intensified</u> <u>allowed Spain to capitalize on the "Special Relationship" between Britain and the United States.</u> ... |
| (INTERPRETATION +) <i>the fact that</i> + EVIDENCE (+INTERPRETATION) |
| ... <u>Domestic interests are reflected in the fact that</u> <u>Putin has recognized that Russia can only be a major player if it has a strong economy.</u> ... |
| ... <i>the fact that</i> <u>relative peace was the result demonstrates that instability is not the inevitable consequence of this course of action</u> ... |

Table 85: *the fact that* used for Cause-Effect and Evidence-Interpretation functions

In two cases in the pooled sample ***the fact that*** was followed by an ‘advantage’ signalled in the element preceding the fixed expression. A similar pattern was seen in two further expressions where ***that*** (ellipted from ***the fact that***) was followed by a ‘reason’ and a ‘problem’ which in each case is signalled in the element preceding it. These patterns are illustrated in Table 86 below.

| |
|---|
| ... saw the 'great advantage ' of managing working-class housing lying in the fact that <u>such buildings 'form[ed] a test-place in which people may prove themselves worthy of higher situations ...</u> |
| ... The principal advantage of a basic income is probably the fact that <u>a guaranteed social wage gives everyone a much larger than normal scope for choosing what to do with their life. ...</u> |
| ... MacPherson's other reason in asserting the 'general rule', that <u>there was a lack of any other suitable term ...</u> |
| ... there are two fundamental problems in giving Ranke the afore-mentioned title - first, that his system was not scientific, and second, that <u>he was not the first to construct such a system ...</u> |

Table 86: the fact that used in expression of advantage, reason and problem

4.5.6 Comparison/contrast across disciplines and institutions

Tables 87, 88 and 89 compare the frequencies of semantic groupings for **that** across the four disciplinary/institutional sub-corpora.

| | HIST-X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|--|--------|--------|--------|--------|----------|--------|----------|--------|
| 'The voice of others' | | | | | | | | |
| AUTHOR OF STATEMENT + CHARACTER OF STATEMENT + <i>that</i> + STATEMENT | 72 | 24.00% | 98 | 32.67% | 72 | 24.00% | 50 | 16.67% |
| CHARACTER OF STATEMENT + <i>that</i> + STATEMENT | 17 | 5.67% | 15 | 5.00% | 32 | 10.67% | 16 | 5.33% |
| CHARACTER OF STATEMENT + IS/WAS + THAT + STATEMENT | 3 | 1.00% | | | 1 | 0.33% | 5 | 1.67% |
| <i>It</i> + CHARACTER OF STATEMENT + <i>that</i> + STATEMENT | 3 | 1.00% | 3 | 1.00% | 2 | 0.67% | 5 | 1.67% |
| TOTAL | 95 | 31.66% | 116 | 38.66% | 107 | 35.66% | 76 | 25.33% |

Table 87: 'The voice of others' – frequency across disciplinary/institutional sub-corpora

| | HIST-X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|--|--------|--------|--------|--------|----------|--------|----------|--------|
| 'The voice of the essay writer' | | | | | | | | |
| 'argument' | | | | | | | | |
| It + ESSAY-WRITER ANGLE ON STATEMENT + <i>that</i> + STATEMENT | 22 | 7.33% | 26 | 8.67% | 21 | 7.00% | 43 | 14.33% |
| I/we + ESSAY-WRITER ANGLE ON STATEMENT + <i>that</i> + STATEMENT | 3 | 1.00% | 3 | 1.00% | 8 | 2.67% | 20 | 6.67% |
| REFERENCE TO ESSAY TEXT ITSELF + <i>will</i> + ESSAY-WRITER ANGLE ON STATEMENT + <i>that</i> + STATEMENT | | | 2 | 0.67% | 5 | 1.67% | 2 | 0.67% |
| TOTAL: 'argument' | 25 | 8.33% | 31 | 10.33% | 34 | 11.33% | 65 | 21.66% |
| 'interpretation' | | | | | | | | |
| PHENOMENON AS EVIDENCE + INTERPRETIVE ANGLE + <i>that</i> + INTERPRETATION | 37 | 12.33% | 31 | 10.33% | 21 | 7.00% | 22 | 7.33% |
| TOTAL: 'argument' and 'interpretaion' | 62 | 20.66% | 62 | 20.66% | 55 | 18.33% | 87 | 29.00% |

Table 88: 'The voice of the essay writer' – frequency across disciplinary/institutional sub-corpora

| | HIST-X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|-----------------------------------|--------|--------|--------|--------|----------|--------|----------|--------|
| PHENOMENON + <i>that</i> + DETAIL | 62 | 20.67% | 54 | 18.00% | 66 | 22.00% | 60 | 20.00% |
| <i>the fact that</i> | 12 | 4.00% | 15 | 5.00% | 19 | 6.33% | 19 | 6.33% |
| Other | 69 | 23.00% | 53 | 17.67% | 53 | 17.67% | 58 | 19.33% |

Table 89: Non ‘voice’ phraseologies for *that* - frequency across sub-corpora

4.5.6.1 Similarities

There are striking similarities across the four disciplinary/institutional sub-corpora samples. In the four sub-corpora ‘voice’ phraseologies (both ‘voice of others’ and ‘voice of the essay writer’) make up between 50 and 60% of the sample, respectively 52.32% in HIST-X, 59.32% in HIST-Y, 53.99% in POL/IR-X and 54.33% in POL/IR-Y. Roughly 20% of the samples for all four sub-corpora consist of PHENOMENON + *that* + DETAIL sequences and *that fact that* sequences constitute between 4% and 6% of all phraseologies across the four sub-corpora.

4.5.6.2 Variations

The clearest differences are neither disciplinary nor institutional. In HIST-X, HIST-Y and POL/IR-X samples there are a higher proportion of ‘the voice of others’ phraseologies than ‘the voice of the essay writer’ phraseologies. For POL/IR-Y this pattern is reversed. The HIST-Y sample contains a noticeably higher number of AUTHOR OF STATEMENT + CHARACTER OF STATEMENT + *that* + STATEMENT sequences, POL/IR-X contains a noticeably higher number of CHARACTER OF STATEMENT + *that* + STATEMENT lines, and, as already mentioned, there are fewer ‘voice of others’ phraseologies in the POL/IR_BAWE sample.

There are no clear disciplinary or institutional patterns in results for ‘the voice of others’, only noticeable idiosyncrasies for particular sub-corpora. Within ‘voice of the essay writer’ the frequency of ‘argument’ phraseologies in POL/IR-Y is approximately double that of the other three sub-corpora. A closer examination of

data in the 300-line random sample for POL/IR-Y reveals that two authors are represented disproportionately in this category, with 38% of instances sourced from essay texts written by 137, whose texts make up only 23% of the sub-corpus. Similarly, 18% of lines in this category come from essay texts written by 234 whose texts only constitute 7% of the sub-corpus.

Three potential disciplinary differences emerged in the first samples: a higher frequency of 'interpretation' within the History sub-corpora, a higher frequency of *I/we* + ESSAY-WRITER ANGLE ON STATEMENT + *that* + STATEMENT sequences within the POL/IR sub-corpora and a small disciplinary difference in the frequency of *the fact that* phraseologies with more instances of these within the POL/IR sub-corpora. However, in each case, these were not confirmed by the second random samples as can be seen in Tables 90 to 92 below.

| <i>I/we</i> + ESSAY-WRITER ANGLE ON STATEMENT + <i>that</i> + STATEMENT | HIST-X | | HIST-Y | | POL/IR - X | | POL/IR -Y | |
|---|--------|-------|--------|-------|------------|-------|-----------|-------|
| Sample 1 | 3 | 1.00% | 3 | 1.00% | 8 | 2.67% | 20 | 6.67% |
| Sample 2 | 1 | 0.33% | 3 | 1.00% | 3 | 1.00% | 5 | 1.66% |

Table 90: Comparison of first and second sample frequencies for *I/we* + ESSAY-WRITER ANGLE ON STATEMENT + *that* + STATEMENT

| PHENOMENON AS EVIDENCE + INTERPRETIVE ANGLE + <i>that</i> + INTERPRETATION | HIST-X | | HIST-Y | | POL/IR - X | | POL/IR -Y | |
|--|--------|--------|--------|--------|------------|-------|-----------|-------|
| Sample 1 | 37 | 12.33% | 31 | 10.33% | 21 | 7.00% | 22 | 7.33% |
| Sample 2 | 31 | 10.33% | 16 | 5.33% | 14 | 4.66% | 20 | 6.66% |

Table 91: Comparison of first and second sample frequencies for PHENOMENON AS EVIDENCE + INTERPRETIVE ANGLE + *that* + INTERPRETATION

| <i>the fact that</i> phraseologies | HIST-X | | HIST-Y | | POL/IR - X | | POL/IR -Y | |
|------------------------------------|--------|-------|--------|-------|------------|-------|-----------|-------|
| Sample 1 | 12 | 4.00% | 15 | 5.00% | 19 | 6.33% | 19 | 6.33% |
| Sample 2 | 5 | 1.66% | 7 | 2.33% | 11 | 3.66% | 7 | 2.33% |

Table 92: Comparison of first and second sample frequencies for the *fact that* phraseologies

4.6. As

4.6.1 Semantic categories for *as*

In the analysis of a 300-line random sample from the pooled corpus containing both 'disciplinary' and institutional sub-corpora, the following categories were identified each of which will be explained in more detail below.

| Category | Number of instances in 300-line sample | Percentage of 300-line sample |
|--|--|-------------------------------|
| 1a. 'Interpretation: perception' | 43 | 14.33% |
| 1b. 'Interpretation: portrayal' | 12 | 4.00% |
| 1c. 'Interpretation: judgement' | 9 | 3.00% |
| 1d. 'Interpretation: classification' | 8 | 2.67% |
| TOTAL for 'Interpretation' | 72 | 24.00% |
| 2. 'Reason - Result' | 33 | 13.33% |
| 3. Fixed expression – <i>such as</i> – 'Exemplification' | 30 | 10.00% |
| 4. Fixed expression – <i>as well as</i> – 'Addition' | 21 | 7.00% |
| 5. 'Attribution' | 21 | 7.00% |
| 6. 'Function' | 22 | 7.33% |
| 7. 'Role' | 17 | 5.67% |
| 8. 'Conceptualisation' | 17 | 5.67% |
| 9. 'Similarity' | 12 | 4.00% |
| 10. 'Measurement/Extent' | 17 | 5.67% |
| 11. 'Simultaneity' | 10 | 3.33% |
| 12. Other | 21 | 7.00% |

Table 93: Semantic analysis of *as* in pooled sample

Again, lines with patterns (including fixed expressions) which occurred very infrequently (i.e. fewer than six times and thus less than 2% of the time) were grouped under 'Other'. The features of each category are discussed below.

4.6.2 'Interpretation'

Just fewer than 25% of the lines in the sample for *as* can be classed as having the semantic motif of 'interpretation'. In lines with this motif the subject or object of interpretation precedes *as* and how this subject or object is interpreted follows *as*. Broadly, with a number of variations in terms of how elements preceding *as* are

ordered, the semantic sequencing for this motif can be described as follows in Table 94:

| |
|--|
| <p align="center">(INTERPRETER[S] &) NATURE OF INTERPRETATION & OBJECT OF INTERPRETATION + <i>as</i> + INTERPRETATION</p> |
|--|

Table 94: Sequencing possibilities for the semantic motif ‘interpretation’

‘*NATURE OF INTERPRETATION*’ refers to the four sub-categories I have divided ‘interpretation’ into, ‘perception’, ‘portrayal’, ‘judgment’ and ‘classification’ (which will be described in more detail below). The three lines in Table 95 below from ‘Interpretation: perception’ provide an indication of the variety of ways in which the elements can be ordered and combined.

| |
|--|
| <p align="center">INTERPRETERS + NATURE OF INTERPRETATION (PERCEPTION) + OBJECT OF INTERPRETATION + <i>as</i> + INTERPRETATION</p> |
| <p align="center"><u>Mao and the politburo</u> <u>considered</u> <u>Taiwan, Vietnam and Korea</u> <u>as</u> <u>potential Cold War arenas</u></p> |
| <p align="center">NATURE OF INTERPRETATION (PERCEPTION) + <i>as</i> + INTERPRETATION + INTERPRETERS + OBJECT OF INTERPRETATION</p> |
| <p align="center"><u>Perceived</u> <u>as</u> <u>a ‘usurper’</u> <u>by</u> <u>Nur al-Din’s suppers</u>, <u>he</u> ...</p> |
| <p align="center">OBJECT OF INTERPRETATION + NATURE OF INTERPRETATION (PERCEPTION) + <i>as</i> + INTERPRETATION</p> |
| <p align="center"><u>they</u> <u>were seen largely</u> <u>as</u> <u>a single commodity</u></p> |

Table 95: An illustration of sequencing possibilities for the semantic motif ‘Interpretation: perception’

The four sub-categories of ‘interpretation’, ‘perception’, ‘classification’, ‘portrayal’ and ‘judgement’ are discussed below in more detail.

4.6.1.1 ‘Interpretation: perception’

In this category the focus is on how someone or something is perceived. Concordance lines for **as** with this semantic motif carry an emphasis on the reception of information by spectators or interpreters. Verbs associated with vision, *see, view, regard, recognise*, and verbs associated with cognition, *understand, perceive, conceive, interpret*, are used in these lines, with forms of the verb *see* appearing in almost half of the instances (see Table 96).

| |
|--|
| ... They <u>saw</u> neurasthenia as a chance to bargain with traditional patriarchy ... |
| ... what was <u>interpreted</u> by the Chinese as fervent American imperialism ... |
| ... <u>the idea of</u> the UN as a stage providing a framework for discussions and multilateral agreement ... |

Table 96: Examples of ‘Interpretation: perception’

4.6.1.2 ‘Interpretation: portrayal’

In this category the focus is on the interpreter’s conveyance of an interpretation to an audience. Thus, in contrast to ‘Interpretation: perception’ the emphasis is productive rather than receptive. Many lines with this motif refer to scholarship in the disciplinary field of focus. The verb *describe* is used in a large proportion of lines carrying this semantic motif (see Table 97).

| |
|--|
| ... All Year Round <u>describes</u> the suburbs as a ‘diluted’ form of London ... |
| ... a feat the anonymous writer <u>describes</u> as a ‘great national sin’. ... |
| ... Orthodox historians, such as Michael Schaller, strongly argue that, as of 1979, Sino-American action cannot be <u>described</u> as a Cold War ... |

Table 97: Examples of ‘Interpretation: portrayal’ (1)

Other verbs used for ‘perception: portrayal’ include *refer to, point to, offer*, and *chart* used when the interpretation involves providing a reason or giving an example as in Table 98 below.

| |
|---|
| ... The “China factor” has been <u>offered</u> as a primary reason for escalation ... |
| ... <u>referring</u> to the witch hunts as drastic examples of mislead [sic] gossip and rumours ... |
| ... They <u>chart</u> its inception as the realisation of “enlightened” principles espoused by the eighteenth-century philosophes ... |

Table 98: Examples of ‘Interpretation: portrayal’ (2)

Lines including the verb *treat* are also included in this category (see Table 99); *treat* tends to be used to convey how a subject or concept is approached within scholarship/analysis, so therefore the framework within which it is portrayed.

| |
|--|
| ... Subjects of politics are <u>treated</u> instrumentally as actors that are unified and alike in their possession of interests and goals ... |
| ... Identity, in its quest for purity <u>treats</u> everything different from it as its polar opposite ... |

Table 99: Examples of ‘Interpretation: portrayal’ (3)

4.6.1.3 ‘Interpretation: judgement’

In this category the focus is on positive or negative interpretation (see Table 100 below). Often these lines have an implied context of debate or disagreement about the object being interpreted. They also often imply a public declaration or display of judgement conveyed by verbs such as *promote*, *hail*, *revere* and *defend* for positive judgement, or *criticise*, *revile*, *denounce* for negative judgement. Some lines in this category could be argued to overlap with the meaning of ‘portrayal’ particularly if they involve a scholarly judgment (see the first two lines in the table below); however, lines with this overlap make up only a small proportion of lines in this category, justifying the broader distinction made here between ‘portrayal’ and ‘judgment’.

| |
|--|
| ... Julian has been <u>reviled</u> by early Christian historians as ‘the very incarnation of evil’ ... |
| ... Bodin <u>defended</u> history as the search for truth ... |
| ... They also isolated themselves by <u>denouncing</u> nationalists as warmongers ... |
| ... it was <u>revered</u> by the Turks as the symbol of religious legitimacy ... |

Table 100: Examples of ‘Interpretation: judgement’

4.6.1.4 ‘Interpretation: classification’

In this category the focus is on how the interpreter is defining, classifying, categorising, grouping or naming the object of interpretation (see Table 101). As with ‘portrayal’, the emphasis is on the productive, rather than receptive, act of locating/positioning the object of interpretation within a context. Verbs such as *class*, *designate*, *identify*, and *define* are typically used in lines with this semantic motif.

| |
|--|
| ... the occupation familiarly <u>designated</u> as ‘shopping’ ... |
| ... He <u>defines</u> master/slave as existent categories within society ... |
| ... the Crusaders were soon <u>identified</u> as formidable fighters with a different agenda to the Greeks ... |

Table 101: Examples of ‘Interpretation: classification’ (1)

This category also takes in an extremely small number of lines containing *know* (as shown in Table 102 below) although this is a slight ‘outlier’ of this grouping; these lines have been grouped in the ‘classification’ rather than ‘perception’ category because they contain the idea of the active ‘labelling’ of a subject or object.

| |
|---|
| ... The East Asian countries which became <u>known</u> as the East Asian “tigers” ... |
|---|

Table 102: Example of ‘Interpretation: classification’ (2)

4.6.3 'Reason-Result'

The second largest grouping of just over 13% of the 300-line random sample can be made under the semantic motif of 'Reason-Result'. In these lines **as** or the fixed expressions **as a result** and **as a result of** create a reason-result link between two propositions. The semantic sequences using **as** that realise this semantic motif are presented in Table 103 below together with examples of concordance lines that follow these patterns.

| RESULT + as + REASON |
|---|
| ... <u>this brought further criticism from the Church</u> as <u>it deviated entirely from church policy</u> ... |
| ... <u>successor states inherited this system from colonial rulers</u> as <u>the elite were indoctrinated before taking power</u> ... |
| ... <u>A lack of belief in the international courts,</u> as <u>they are essentially flawed,</u> ... |
| as + REASON + RESULT |
| ... Perhaps as <u>it was a question of morality in the public sphere</u> <u>it leant more easily to the opinion of women even though it was in the political sphere.</u> ... |

Table 103: 'Reason-Result' semantic sequences with **as**

The semantic sequences with either **as a result** or **as a result of** (N.B. *consequence* can also be used in place of **result**) that realise this semantic motif within the 300-line pooled sample are presented in Table 104 below, again, together with examples of concordance lines that follow these patterns.

| |
|---|
| RESULT + <i>as a result of</i> + REASON |
| <p>... <u>associations have developed</u> <i>as a result of</i> <u>the course taken by David</u> ...</p> <p>... his theory of <u>exclusion from the franchise</u> <i>as a result of</i> <u>the dependence on the will of other men</u> ...</p> <p>... <u>will most likely lead to an increase in losses</u> <i>as a result of</i> <u>bargaining between the two poles of power.</u> ...</p> |
| <i>as a result of</i> + REASON + RESULT |
| <p>... The intervening time, <i>as a result of</i> the <u>policies deployed by the military during that time</u>, has <u>robbed Manning of the respect connected to his position as Private fc</u>, ...</p> |
| REASON + RESULT + <i>as a result</i> |
| <p>... <u>environmental policy will struggle to progress as it will become somewhat re-nationalised</u> and <u>global protection may suffer</u> <i>as a result</i>. ...</p> |

Table 104: 'Reason-result' semantic sequences with *as a result* or *as a result of*

4.6.4 Fixed expression – *such as* – 'Exemplification'

The most frequently found fixed expression is *such as* making up 10% of the random sample and always used to list examples of a phenomenon directly preceding the expression (see Table 105 below).

| |
|---|
| PHENOMENON + <i>such as</i> +EXAMPLE[S] |
| <p>... <u>traditional 'Cold War' manifestations</u> <i>such as</i> <u>nuclear arms races and internationalised conflicts</u> ...</p> <p>... <u>seventeenth-century personalities</u>, <i>such as</i> <u>John Reeve</u> and <u>Lodowick Muggleton</u>, ...</p> <p>... <u>Iraq's oil reserves</u> <i>such as</i> <u>those of Mosul or Kirkuk</u> ...</p> <p>... <u>human rights abuses</u> <i>such as</i> <u>in the Tiananmen Square Massacre of 1989</u> ...</p> |

Table 105: 'Exemplification' with *such as*

4.6.5. Fixed expression – *as well as* – ‘Addition’

The second most frequently found fixed expression in the random sample, occurring in just under 7% of lines, was *as well as* used to add either another subject/object or action. Having attempted to divide up the twenty lines into sub-categories of the ‘addition’ semantic motif (with potential semantic elements including **ACTOR**, **GROUP**, **ACTION**, **INTENTION**, **SENSE** and **STATE**), I made the decision this was going one level of categorisation delicacy too far in this case as the resulting sub-categories would each contain between one and three lines. I therefore decided to use **PHENOMENON** in its broadest sense to label the semantic elements preceding and following *as well as*. The two semantic sequences for this motif are illustrated in Table 106 below.

| PHENOMENON[S] + <i>as well as</i> + PHENOMENON[S] |
|--|
| <p>... <u>the relationship between the structure and the agency, as well as the construction of state and institutional interests</u> ...</p> <p>... <u>economics, distribution, housing, labour protection and social security as well as public health</u> ...</p> <p>... y’ or <u>to merely ensure cooperation to a limited extent between its members as well as to temper inter-Arab suspicions</u> ...</p> |
| <i>as well as</i> + PHENOMENON[S] + PHENOMENON[S] |
| <p>... <i>As well as</i> <u>anti-Semitism</u>, the paper contained <u>populist economic and foreign policy messages</u>. ...</p> <p>... <i>As well as</i> <u>someone being able to decide what they want to do with their life a UBI gives people the choice of when to do certain thing</u> ...</p> |

Table 106: Semantic sequences with *as well as* for ‘addition’

4.6.6 ‘Attribution’

The semantic motif of this grouping is that of attributing a proposition, often in the form of a direct quotation, to an author (see Table 107 below). In just under three quarters of these lines the word *as* begins the sequence.

| as + ATTRIBUTION TO AUTHOR + PROPOSITION |
|---|
| <p>... As Weber acknowledges, <u>the East possessed geometry, natural sciences, and chemistry, but what was lacking was a method of experiment and rationality.</u> ...</p> <p>... As Jack Goldsmith writes, <u>'[t]he ICC is an intricate, and in some respects impressive, legal edifice. But a legal edifice devoted to international peace and stability is worthless without some plausible mechanism of enforcement'.</u> ...</p> <p>... but as Peter Reil has argued, <u>his spontaneity is distinct from the mechanistic notions one associates with sciences.</u> ...</p> <p>... Furthermore, as Newman puts it, <u>identity has a 'discursively constituted nature'</u> ...</p> |
| PROPOSITION + as + ATTRIBUTION TO AUTHOR |
| <p>... If <u>seclusion was necessary and doctor-patient relations vital</u> as Schuster <u>illustrates,</u> ...</p> |
| as + EXPLICIT EVALUATIVE ATTRIBUTION TO AUTHOR + PROPOSITION |
| <p>... As Carr convincingly argues, <u>a 'scientific' method of history is not enough to render history a science in itself.</u> ...</p> <p>... As Vainker astutely points out: <u>despite the high quality of silk, porcelain and lacquer, these goods were not perceived as luxury items, but as functional goods by the affluent Chinese.</u> ...</p> |

Table 107: Semantic sequences for 'Attribution'

A range of reporting verbs are deployed across these twenty lines, with *argue* (used four times), *point out* (used three times), and *state* (used twice), being the only verbs occurring more than once. When an adverb has been added which evaluates the reported proposition - in this sample this occurs three times with adverbs used being *convincingly*, *astutely* and *rightly* – the line has been labelled under a second semantic sequence with the element **EXPLICIT EVALUATIVE ATTRIBUTION TO AUTHOR**. Obviously, the reporting verb itself in an attributive phrase can carry an implicit evaluative meaning, but, considering the small size of the 'attribution' category, further subdivision along these lines is not useful.

4.6.7 'Function'

Another grouping to be found within the sample is that of lines which carry the semantic motif of 'function'. In each case the head noun from the phrase following *as* collocates quite naturally with the verbs *use*, *serve*, or *function*.

This is illustrated in Table 108 below which lists all nouns from this category designated as 'head noun' in the phrase following *as* (N.B. the verbs *use* and *serve* preceding *as* are actually contained in a small but significant number of these lines).

| | | |
|-----------------------------|----|--|
| used serves functions | as | a laboratory a forum a 'goodwill gesture' a source an illustration a guide a medium a court a gateway a tool a 'building bloc' an alternative |
|-----------------------------|----|--|

Table 108: Head nouns' following *as* in 'Function' phraseologies

One line in the sample does not contain a head noun following *as* which collocates well with the verbs listed (see Table 109 below). However, when examined, the meaning of the head noun *bastion* in the context of this line is actually that of *source*, a word which does collocate with the listed verbs.

| |
|--|
| ... remains as a <u>bastion</u> from which the critics of luxury draw their ammunition ... |
|--|

Table 109: *bastion* used as *source*

The semantic sequencing of this semantic motif is quite flexible (see Table 110). In some cases an actor, 'utiliser' of the tool and a verb describing the way in which the tool is utilised are present, but these are necessary conditions for a phrase being included within this grouping. A phrase is categorised as part of this grouping if the phenomenon preceding *as* is framed as a function by the language following *as*.

(UTILISER) & (PROCESS OF UTILISATION) + OBJECT BEING UTILISED + *as* + FUNCTION

Table 110: Sequencing possibilities for the semantic motif ‘Function’

These sequences fell into three sub-groupings identified and exemplified in Table 111 below.

| OBJECT BEING UTILISED + <i>as</i> + FUNCTION |
|---|
| ... <u>the unified individual as key ‘building-bloc’ of any social theory of agency and culture</u> ... |
| ... introduced <u>the seminar as a laboratory to teach engagement with sources</u> ... |
| OBJECT BEING UTILISED + PROCESS OF UTILISATION + <i>as</i> + FUNCTION |
| ... <u>It has been little used as a source of evidence for Leveller attitudes</u> ... |
| ... <u>pieces by other artists used as additional illustration</u> ... |
| <i>as</i> + FUNCTION + OBJECT BEING UTILISED |
| ... <u>as the “gateway to the Middle East”, the Mediterranean</u> ... |

Table 111: Semantic sequences for ‘Function’

4.6.8 ‘Role’

This category has features in common with the category of ‘function’; ‘role’ can be seen as the ‘function’ of ‘an actor’ or ‘actors’ and the sequencing possibilities for this motif are illustrated in Table 112 below.

(ROLE-PLAYER) OR (ROLE-PLAYER’s *role/status/image* etc.) + *as* + ROLE + (ROLE-PLAYER)

Table 112: Sequencing possibilities for the semantic motif ‘Role’

The possessor of the ‘role’ in these phrases is usually a human subject or object but sometimes also a nation or entity in anthropomorphised form as in the following two examples in Table 113 containing *Germany* and *NATO* respectively.

| |
|---|
| ... liquidating <u>Germany</u> as <u>an imperialist competitor</u> ... |
| ... The latter missions are part of NATO's new role as <u>a risk manager</u> |

Table 113: Examples of 'Role' (1)

There are also some instances in which there is arguably an overlap of 'role' and 'function' where the human subject/object in the concordance line is passively used as a tool by another human actor (see Table 114 below). In all these cases, concordance lines are categorised under 'role' because of the presence of a human actor.

| |
|--|
| ... his use of <u>Petty</u> as <u>representative for the rest of the movement</u> ... |
| ... in their historiography <u>Ranke</u> was more important as <u>a symbol</u> rather than as a historian ... |

Table 114: Examples of 'Role' (2)

Concordance lines in the category of 'role' take the following patterns set out in Table 115, the vast majority taking the first pattern listed.

| |
|---|
| as + ROLE |
| ... Ranke's image as <u>'the father of scientific history'</u> his definitive status as <u>sacrosanct party leader at the Tenth Congress</u> her role as <u>the symbolic intimate</u> prominent as <u>a writer</u> ... |
| as + ROLE + ROLE-PLAYER |
| ... even as <u>an ardent pagan</u> <u>Ammanianus</u> is willing to ... |

Table 115: Semantic sequences for 'Role'

4.6.9 'Conceptualisation'

In this category the phrase following **as** expresses a particular conceptualisation (which can neither be defined as 'function' nor 'role') of the phenomenon in the phrase (usually) preceding **as** (see Table 116 below for sequencing possibilities).

PHENOMENON + (STATE OR PROCESS) + *as* + CONCEPTUALISATION

Table 116: Sequencing possibilities for the semantic motif ‘Framing or Conceptualisation’

The different sequential variations and examples of each are shown in Figure 4.80 below.

| |
|--|
| PHENOMENON + <i>as</i> + CONCEPTUALISATION |
| ... <u>normality</u> <i>as</i> <u>peace</u> ... |
| PHENOMENON + PROCESS + <i>as</i> + CONCEPTUALISATION |
| ... which saw the <u>Internet</u> <u>develop</u> <i>as</i> <u>a cold war project</u> ... |
| ... <u>the fight</u> <u>occurred</u> <i>as</i> <u>a pseudo-historical event</u> ... |
| PHENOMENON + STATE + <i>as</i> + CONCEPTUALISATION |
| ... <u>its existence</u> <i>as</i> <u>a constitutive part of our identity</u> ... |
| <i>as</i> + CONCEPTUALISATION + PHENOMENON |
| ... and <i>as</i> <u>a concept</u> <u>it</u> conjured up an abundance of political whitewash for him ... |

Figure 4.80: Semantic sequences for ‘Framing or Conceptualisation’

4.6.10 ‘Similarity’

The semantic motif linking concordance lines in this category is similarity or correspondence between two phenomena with *as* carrying the meaning of ‘in the way’, or, more colloquially, ‘like’ (see Table 117).

| |
|--|
| PHENOMENON + <i>as</i> + SIMILAR PHENOMENON |
| ... One cannot <u>understand the new science of history</u> <i>as</i> <u>it was understood by Ranke</u> without ... |
| ... <u>too much Hellenism</u> in Arnolds [sic] eyes would <u>spell the end for a state</u> just <i>as</i> <u>Hebraism was stifling England</u> |
| ... hat <u>differed from state to state and that had no pretension to imperialism</u> <i>as</i> <u>was</u> <u>the case for Iraq</u> ... |
| ... <u>the development of a collective identity</u> , <i>as</i> <u>observed in the Zapatista case</u> , ... |
| ... <u>the constricting nature of the term genocide</u> , <i>as</i> <u>outlines[sic] in the Genocide</u> <u>Convention</u> ... |

Table 117: ‘Similarity’ sequences

Two lines grouped in this category from the pooled sample ostensibly have features that could make them members of the ‘attribution’ category (see Table 118). These lines have been categorised as ‘similarity’ rather than ‘attribution’ phraseologies because the *as* in each has been judged to be emphasising the similarity or correspondence between two ideas rather than merely expressing the attribution of a statement to a particular author. Again, this provides a useful illustration of the ‘overlaps’ or ambiguities that can occur in this process of categorisation which require decisions to be made which are very much qualitative.

| |
|---|
| <p>... <i>As</i> I said above, I am not concerned with the case of the family with young ...</p> <p>... Third, just <i>as</i> Arthur Balfour argued in 1910 that the Orient "...exists in a sense, as ...</p> |
|---|

Table 118: ‘Ambiguous’ lines which are categorised as ‘similarity’ rather than ‘attribution’

4.6.11 ‘Emphasis of extent’

Lines in this category contain an ‘as ... as’ pattern or a variation of it which on the surface may suggest a semantic motif of equality/sameness. In fact, the ‘*as ... as*’ pattern seems generally to be used to emphasise or intensify a measurement or judgement of a particular situation or phenomenon.

| (STATE OR PROCESS) + <i>as</i> + MEASUREMENT OF EXTENT + <i>as</i> + VALUE BY WHICH EXTENT IS BEING MEASURED |
|--|
| <p>... <u>As far back as 1977</u>, ...</p> <p>... <u>exploiting as much as they could</u> ...</p> <p>... nothing <u>represented a sacrifice on behalf of the nation as significantly as the actions of his real contemporaries</u>. ...</p> <p>... as private contractors sought to ‘<u>throw up</u>’ <u>as many districts as required by the steady increase of outer London’s popu</u> ...</p> <p>... he <u>is as little free to have it as he would be if it were forbidden him by law</u>. ...</p> |

Table 119: Semantic sequences for ‘Emphasis of extent’

In the first example in Table 119 above, the ‘*as ... as*’ pattern is used to emphasise the earliness of events being discussed; it functions to express this timing, 1977, as unusually or unexpectedly early. In the second example the ‘*as ... as*’ pattern functions to emphasise/intensify the extent of exploitation. Even in the line in Table 120 below, which could be argued to on the surface be expressing equality/sameness between two phenomena, the important function of ‘*as ... as*’ is in fact emphasis/intensification. The ‘*as ... as*’ pattern here serves to emphasise the fact that so many non-historians contributed rather than the fact that an equal number of historians did so; just as ‘*As far back as 1977*’ expresses unexpected/unusual earliness, ‘*as much as historians*’ expresses as surprising/unexpected the level of contribution to the Annales, a history, from non-historians.

| |
|---|
| <p>... the Annales have <u>drawn their inspiration</u> from <u>non-historians</u> <i>as much as</i> <u>historians</u> ...</p> |
|---|

Table 120: Emphasis/intensification realised with *as much as*

4.6.12 ‘Simultaneity’

In this category *as* shows the relationship of simultaneity between two events/actions and is realised through the two sequences illustrated in Table 121 below.

| |
|---|
| ONE OF TWO SIMULTANEOUS EVENTS/ACTIONS + <i>as</i> + THE OTHER OF TWO SIMULTANEOUS EVENTS/ACTIONS |
| ... <u>and was, even then, barbarised</u> <i>as</i> <u>the Franks developed their own literary style and culture.</u> ... |
| ... <u>Czech cultural integrity faded</u> <i>as</i> <u>the fog of war descended over Europe</u> ... |
| <i>as</i> + ONE OF TWO SIMULTANEOUS EVENTS/ACTIONS + THE OTHER OF TWO SIMULTANEOUS EVENTS/ACTIONS |
| ... <i>as</i> <u>the supply of medical services increased,</u> <u>demand fell rapidly,</u> ... |

Table 121: Semantic sequences for 'Simultaneity'

4.6.13 Comparison/contrast across disciplinary and institutional sub-corpora

The results of this analysis of the disciplinary and institutional sub-corpora are presented in Table 122 below. No new semantic categories were identified in this process.

| Category | HIST-X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|---------------------------------------|-----------|---------------|-----------|------------|-----------|---------------|-----------|---------------|
| 1a. 'Interpretation: perception' | 32 | 10.67% | 50 | 16.67% | 48 | 16.00% | 27 | 9.00% |
| 1b. 'Interpretation: portrayal' | 18 | 6.00% | 14 | 4.67% | 16 | 5.33% | 12 | 4.00% |
| 1c. 'Interpretation: judgement' | 7 | 2.33% | 14 | 4.67% | 3 | 1.00% | 6 | 2.00% |
| 1d. 'Interpretation: classification' | 8 | 2.67% | 3 | 1.00% | 6 | 2.00% | 2 | 0.67% |
| TOTAL for 'Interpretation' | 65 | 21.66% | 81 | 27% | 73 | 24.33% | 47 | 15.66% |
| 2. 'Reason - Result' | 37 | 12.33% | 20 | 6.67% | 39 | 13.00% | 69 | 23.00% |
| 3. 'Exemplification' – <i>such as</i> | 51 | 17.00% | 15 | 5.00% | 33 | 11.00% | 21 | 7.00% |
| 4. 'Addition' – <i>as well as</i> | 19 | 6.33% | 9 | 3.00% | 25 | 8.33% | 24 | 8.00% |
| 5. 'Attribution' | 3 | 1.00% | 22 | 7.33% | 23 | 7.67% | 12 | 4.00% |
| 6. 'Function' | 19 | 6.33% | 28 | 9.33% | 22 | 7.33% | 14 | 4.67% |
| 7. 'Role' | 21 | 7.00% | 31 | 10.33% | 14 | 4.67% | 21 | 7.00% |
| 8. 'Conceptualisation' | 7 | 2.33% | 16 | 5.33% | 15 | 5.00% | 21 | 7.00% |
| 9. 'Similarity' | 17 | 5.67% | 26 | 8.67% | 16 | 5.33% | 15 | 5.00% |
| 10. 'Measurement/Extent' | 21 | 7.00% | 16 | 5.33% | 17 | 5.67% | 23 | 7.67% |
| 11. 'Simultaneity' | 12 | 4.00% | 17 | 5.67% | 6 | 2.00% | 10 | 3.33% |
| 12. Other | 24 | 8.00% | 15 | 5.00% | 14 | 4.67% | 18 | 6.00% |

Table 122: Frequency for semantic groupings with *as* across disciplinary/institutional sub-corpora

4.6.13.1 Similarities

This analysis has shown broad patterns of similarity rather than patterns of difference to be most clearly identifiable. Across all four corpora there is a marked preference for use of **as** for the purposes of ‘interpretation’, making up between 15% and 27% of the sub-corpora samples, with the majority of these uses being to express ‘perception’, which make up between 9% and 17% of the sub-corpora samples. There is also broad similarity across all four corpora in terms of the proportion of uses for ‘function’, ‘role’ and ‘quality/characteristic’, and for ‘similarity, ‘extent’ and ‘simultaneity’.

4.6.13.2 Variations

There are no patterns that can confidently be identified as attributable to difference based on institution or discipline. Instead, there are categories for which the results from one sub-corpus is noticeably ‘skewed’ from the average: POL/IR-Y has a noticeably smaller number of lines used for ‘interpretation’ and a significantly larger number of uses for ‘reason’; HIST-X has a noticeably larger number of lines containing the fixed expression ‘such as’.

When examined more closely, both these results can be attributed to the influence of particular individual authors within each sub-corpus. In the latter case, 40% of the ‘such as’ lines were from one student within the HIST-X sub-corpora; this student’s six essays make up roughly 20% of the sub-corpus, so this represents quite a substantial over-contribution to that particular meaning within the sample.

For the POL/IR-Y sub-corpora the result can also be argued to be due to individual writer influence with two students contributing over a third each (36%) to the ‘reason’ category of lines within the sample. However, because of the small size and limited number of contributors to this sub-corpus, it cannot be argued that this is a particularly unusually high contribution from these students to this particular

meaning as both students have each contributed just under a third of the essays (31%) that make up the sub-corpus.

4.7 This

Before discussing the semantic analysis of **this** within the current study, differences between how I have approached the keyword in comparison to how Groom approached a similar keyword in his study will be discussed. Although in Groom's (2007) study **this** was not a keyword, its plural form, **these**, a word which has similar properties to **this**, did appear. The approach in the current study to the analysis of **this** differs in important ways to how Groom treated **these** within his analysis. Groom's treatment of **these** was briefly outlined in the Methodology chapter, but a more detailed explication of differences in approach is useful at this stage.

4.7.1 Comparison with Groom's (2007) approach

As briefly outlined in section 3.7 of the Methodology Chapter, Groom divides his concordance lines grammatically into '**these n**' in which **these** functions as a determiner and '**these as pronoun (anaphoric)**'. Nothing more is done with the latter category which accounted for a small 10% and 6% of the HistArt and LitArt corpora samples respectively.

Groom expands the co-text in the concordance lines for **these n** phrases and categorises the sequences by analysing the 'referents', the longer statement usually preceding but sometimes following **these**, to which the '**these** + noun phrase' sequences were referring. In this way, he firstly categorises sequences according to Sinclair's (2004) taxonomy as either *encapsulating*, 'the process of simultaneously condensing the content (or part of the content) of a previous statement and recycling it as the starting point of a new proposition' (Groom, p. 186), or

prospecting, signalling ‘what will come next or at some later point in the text’ (pp. 185-186).

Groom then sub-categorises the much larger *encapsulation* grouping under either ‘summary’, summarising the preceding statement, or ‘repetition’ repeating a word which is part of the preceding statement. The larger ‘summary’ category is then divided according to whether the summaries were ‘neutral’ (e.g. *these proposals* (p. 186)) or ‘evaluative’ (e.g. *these fanciful etymologies* (ibid.)). A reproduction of the resulting table from Groom’s study is shown in Figure 5 below.

| | | | | HistArt | LitArt |
|------------------------------|---------------|------------|------------|---------|--------|
| these n | | | | 90 | 94 |
| | Encapsulation | | | 90 | 89 |
| | | summary | | 59 | 66 |
| | | | neutral | 56 | 58 |
| | | | evaluative | 3 | 8 |
| | | repetition | | 31 | 23 |
| | Prospection | | | 0 | 5 |
| These as pronoun (anaphoric) | | | | 10 | 6 |

Figure 5: Reproduction of Groom’s (2007, p. 187) ‘Table 5:19: Grammatical and semantic roles of *these* in HistArt and LitArt’

Arguably, the approach Groom takes to analysis of ***these*** departs somewhat in spirit from the stated purpose of his approach, namely the bottom-up semantic rather than grammatical analysis of phraseologies within which target keywords occur. Firstly, by adopting Sinclair’s (2004) pre-existing framework, the analysis of ***these*** is no longer a purely bottom-up exercise. Secondly, within this framework categorisation of ***these*** phraseologies is primarily based on analysis of the referents within the co-text rather than the meaning within phraseologies themselves. The subsequent divisions, encapsulation, prospection, summary and repetition, are more functional than semantic. The only truly semantic division made is the very broad-brush distinction between ‘neutral’ and ‘evaluative’ summaries.

My approach to analysis of ***this*** differs from Groom’s approach to analysis of ***these*** in a number of important ways. Firstly, rather than go directly to the referent co-text to deduce meaning, semantic properties and groupings are primarily

determined by focusing on **this** phraseologies themselves, i.e. the semantic sequences which **this** occurs *within*. Co-text is referred to in cases where meaning within the phraseology is ambiguous, but the aim is for semantic grouping to be determined wholly or mainly from the phraseologies themselves. Secondly, the semantic divisions of **this** phraseologies within the present study are more fine-grained than the broad-brushed ‘neutral’ versus ‘evaluative’ groupings for the **these** phraseologies in Groom’s study. Thirdly, my analysis looks not only at the semantic properties of ‘**this** + noun/noun phrase’ sequences, but also at sequences in which **this** operates grammatically as a pronoun. The equivalent sequences in Groom’s study, sequences in which **these** operated as a pronoun, were not analysed for semantic properties. In my pooled sample sequences in which **this** operates as a pronoun make up approximately 37 percent of all sequences. Thus, the meanings associated with these sequences are an important component of its use in the student texts.

It needs be noted that the two major grammatical patterns for **this** - its use as a determiner and its use as a pronoun - do strongly link to how phraseologies are semantically grouped: categories 1 to 7 outlined below share the idea of ‘encapsulation’ (Sinclair, 2004; Groom, 2007) associated with ‘**this** + noun/noun phrase’ grammatical patterns, and categories 8-14, all of which carry semantic motifs to do with comment of some kind on the information in the text which precedes **this**, are associated with the use of **this** as a pronoun carrying anaphoric reference.

4.7.2 Semantic categories for **this**

Fifteen separate semantic groupings were identified for **this** from the pooled sample (numbers 1-15 in Table 123). A sixteenth was added having been found to occur more than six times in the POL/IR-Y sub-corpora. As discussed above, for categories 1 to 7 I have, as a starting point, appropriated the concept

‘encapsulation’ from both Sinclair (2014) and Groom (2007) used here to describe the condensing and summary *or* repetition of ideas in the text preceding **this** and the semantic sequence occurring with **this**.

| Category | Number of instances in 300-line sample | Percentage of 300-line sample |
|---|--|-------------------------------|
| 1. ‘Encapsulation as scholarly angle’ | 31 | 10.33% |
| 3. ‘Encapsulation as object’ | 30 | 10.00% |
| 2. ‘Encapsulation as process’ | 29 | 9.67% |
| 4. ‘Encapsulation of the essay text itself’ | 20 | 6.67% |
| 5. ‘Encapsulation as situation or circumstance’ | 16 | 5.33% |
| 6. ‘Encapsulation as sense’ | 11 | 3.67% |
| 8. ‘Evaluation of preceding statement’ | 38 | 12.67% |
| 9. ‘Interpretation of preceding statement’ | 22 | 7.33% |
| 10. ‘Effect of preceding statement’ | 17 | 5.67% |
| 11. ‘Addition to preceding statement’ | 17 | 5.67% |
| 12. ‘Exemplification of preceding statement’ | 9 | 3.00% |
| 13. ‘Cause of preceding statement’ | 7 | 2.33% |
| 14. Semi-fixed expression - ‘Point in time’ | 7 | 2.33% |
| 15. Semi-fixed expression – <i>in this manner/respect/sense/way</i> | 4 | 1.33% |
| 16. Other | 42 | 14.00% |

Table 123: Semantic analysis of **this** in pooled sample

4.7.3 ‘Encapsulation’

As stated in section 4.4.1 above the concept of ‘encapsulation’ is appropriated from Sinclair (2004) and Groom (2007) to describe the summary or repetition of an idea occurring in the text preceding **this** occurring within the noun or noun phrase that follows **this**. (its meaning is arguably stretched for category four, ‘encapsulation of the essay text’, as these sequences entail simultaneous anaphoric, *encapsulating*, and cataphoric, *prospecting*, reference).

As also stated in section 4.4.1, the co-text has not routinely been used to determine what kind of encapsulation is happening within the **this** sequences. Instead, as far

as possible, sequences have been judged on their internal logic with regards to how ideas in the preceding text are being interpreted within their summary or repetition in these sequences.

Because of the diverse range of types of meaning in the noun and noun phrase following **this** in these phraseologies, the semantic grouping of ‘encapsulation’ sequences posed a particular challenge with regards to level of delicacy. For these phraseologies, a judgement needed to be made regarding ‘cut off’ for delicacy which meant the resulting groupings were neither too small nor too wide-ranging as to make them unwieldy. This ‘cut off’ is inevitably somewhat arbitrary entailing compromises reflected in ambiguities at the less proto-typical ‘borders’ of the resulting groupings. These ambiguities/overlaps will be discussed as part of the analysis below.

Six broad categories were settled upon for the **this** + noun/noun phrase phraseologies; ‘encapsulation as scholarly angle’, ‘encapsulation as process’, ‘encapsulation as object’, ‘encapsulation as the essay text’, ‘encapsulation as circumstance’, and ‘encapsulation as sense’.

4.7.3.1 ‘Encapsulation as scholarly angle’

This is the largest ‘encapsulation’ grouping in the pooled 300-line sample. In this grouping the noun or noun phrase following **this** encapsulates the preceding statement as a particular scholarly way of approaching, framing or interpreting an aspect the field or phenomena with which the essay is concerned. Sequences grouped under this motif contain the element **CONCEPTUAL LABELLING** discussed in sub-section 4.2.2.8 of the analysis of keyword **of**.

In most cases this motif is expressed in the semantic sequence **this** + **CONCEPTUAL LABELLING** with the ‘labels’ including the single words *distinction*, *point*, *notion*,

question, history and theory or phrases such as *conceptual lens* and *theoretical framework* (illustrated in Table 124 below).

| <i>this</i> + CONCEPTUAL LABELLING |
|---|
| ... this <u>distinction</u> ... |
| ... this <u>interpretation</u> ... |
| ... this <u>notion</u> ... |
| ... this <u>question</u> ... |
| ... this <u>theory</u> ... |

Table 124: 'Encapsulation as scholarly angle' (1)

In some cases an evaluation of the labelling is included creating the semantic sequence ***this* + DESCRIPTION OR EVALUATION + CONCEPTUAL LABELLING** (illustrated in Table 125 below).

| <i>this</i> + DESCRIPTION OR EVALUATION + CONCEPTUAL LABELLING |
|---|
| ... this <u>abstract argument</u> ... |
| ... this <u>relativist approach</u> ... |

Table 125: 'Encapsulation as scholarly angle' (2)

Four lines in the pooled random sample have an additional semantic element. Two contain the particular concept a 'label' applies to creating the semantic sequence ***this* + (DESCRIPTION OR EVALUATION) + CONCEPTUAL LABELLING + *of* + CONCEPT** sequences in Table 126 below. Notice this sequence contains the semantic sequence **CONCEPTUAL LABELLING + *of* + CONCEPT** for the keyword *of*.

| <i>this</i> + (DESCRIPTION OR EVALUATION) + CONCEPTUAL LABELLING + <i>of</i> + CONCEPT |
|---|
| ... this <u>construct of the stereotyped enemy</u> ... |
| ... this <u>reified theory of "collective identity"</u> ... |

Table 126: 'Encapsulation as scholarly angle' (3)

Two others in Table 127 below contain the object of conceptual labelling, creating the semantic sequence **this** + CONCEPTUAL LABELLING + *of/between* + OBJECT sequences (notice that in the second line ‘this view of Ranke’ would be categorised as **PROCESS** + **of** + OBJECT if it were an ‘*of*’ line).

| this + CONCEPTUAL LABELLING + <i>of/between</i> + OBJECT |
|---|
| ... this <u>distinction between them</u> ... |
| ... this <u>view of Ranke</u> ... |

Table 127: ‘Encapsulation as scholarly angle’ (4)

4.7.3.2 ‘Encapsulation as Object’

Semantic sequences with **this** in this grouping encapsulate an ‘object’ ranging from more concrete instances such as geographical areas (*British outpost, region*), people (*extraordinary woman, healer*), and objects (*painting*), to the abstract *faction, term, aspect of the Annales, highest level of Hegemony* and *type of fiction* (see Table 128 below).

| this + OBJECT |
|---|
| ... this <u>aspect of the Annales</u> ... |
| ... this <u>extraordinary woman</u> ... |
| ... this <u>“paper tiger”</u> ... |
| ... this <u>primarily economic work</u> ... |
| ... this <u>term</u> ... |

Table 128: ‘Encapsulation as object’ (1)

A noticeable subset involves ‘texts’ of various kinds, *article, piece, play, primarily economic work, source* and *text*. Another noticeable subset involves ‘method’ nouns (see Table 129 below). These could possibly be interpreted as ‘processes’ in that they entail ‘ways of doing things’; however, they do not denote a particular instance of an event but rather a *set of principles* for a process rather than the process itself and have therefore been categorised as abstract objects. Some lines including the

word *approach* are grouped here rather than in ‘encapsulation of scholarly angle’ when it is not describing a scholarly method but instead an approach of a particular actor or actors within the field which is the subject of the essay.

| |
|---|
| ... this <u>approach</u> ... |
| ... this <u>scheme</u> ... |
| ... this <u>system</u> ... |
| ... this <u>surveillance 2.0</u> ... |

Table 129: ‘Encapsulation as object’ (2)

4.7.3.3 ‘Encapsulation as Process’

In this grouping the statement preceding **this** is encapsulated as a ‘process’. Nouns in the semantic sequences which begin with **this** illustrated in Table 130 below denote events, e.g. *interplay*, *trade*, *shift*, *transcendence*, *crisis* and *warring*, or experiences, e.g. *hardship* and *experience*.

| this + PROCESS |
|--|
| ... this <u>interplay</u> ... |
| ... this <u>luxury trade</u> ... |
| ... This <u>shift in population and employment patterns</u> ... |
| ... this <u>hardship</u> ... |
| ... This <u>negative experience</u> ... |

Table 130: ‘Encapsulation as process’

The majority of sequences in this grouping were easily categorised as ‘encapsulation of process’ sequences based on the nature of the main noun following **this**. In a few cases categorisation entailed a closer examination of co-text. Table 131 below shows the co-text for three sequences containing the nouns *discipline*, *phenomenon* and *line* respectively, which were at first intuitively assigned to this grouping. Examination of the co-text both preceding and following the semantic sequence proves initial intuitions correct: all three sequences are encapsulating a process.

... The intervening time, as a result of the policies deployed by the military during that time, has robbed Manning of the respect connected to his position as Private 1st class, the social capital attached to his considerable intelligence, his dignity and agency as an individual human being. The effectiveness of this discipline is not aleatory. Every step in the management of Manning's detainment, from the length of the confinement, the place of confinement and the style of confinement, is the product of decisions made using a specific source of knowledge - that of military techniques of discipline related to breaches of this kind. ...

... It is claimed that it is in the interest of political leaders to keep tribal rivalry and ethnic violence as a part of national politics, as they can use this to their advantage in asserting their power. Again this phenomenon is prevalent in Kenya as political parties, such as Kenya African National Union (KANU), who are known to be affiliated to a particular ethnic group, put their tribes' interests above national interests and they use violence as a means to gain control over other ethnic groups and assert their authority....

... Hill fervently believed that 'you cannot deal with the people and their houses separately' as only through the moral 'disciplining of our immense poor population' could London's 'mob of paupers [be changed] into a body of self-dependent workers'. Indeed, this strict moral line can be traced throughout Hill's long career as a pioneer of social housing, as whilst she strove to ensure that houses under her charge remained 'in a tenable order', despairing at the 'deplorable condition' which her fellow landlords kept their buildings in, she did so in such a way as to 'enforce right' and 'cast out the sin' from the lives of her tenants. ...

Table 131: Examination of co-text to confirm membership within grouping of problematic 'process' lines

4.7.3.4 'Encapsulation of the essay text itself'

This is a clear and simple category (see Table 132 below); there are no ambiguities with regard to which sequences should be grouped here. In all instances the noun following **this** identifies either the whole (*essay, paper, enquiry*) or part (*section*) of the essay in which the sequence appears. Of the twenty sequences carrying this meaning, twelve contain the noun *essay*, six *paper* and there is once instance each of *enquiry* and *section*.

... This enquiry ...
 ... This essay ...
 ... this paper ...
 ... this section ...

Table 132: *this* + PART/WHOLE OF ESSAY TEXT ITSELF

4.7.3.5 'Encapsulation *as* situation or circumstance'

In this grouping illustrated in Table 133 below it is not actions or events that are encapsulated but the situations, circumstances and states of affairs that often serve as the background or context within which events, actions or experiences take place. This category includes sequences containing nouns such as *tension*, *friction* and *imbalance* which describe an ongoing 'climate' or set of circumstances rather than an instance of a particular process.

... this context ...
 ... this culture ...
 ... This homogeneity ...
 ... this imbalance ...
 ... this potential for a brighter future ...

Table 133: *this* + SITUATION OR CIRCUMSTANCE (1)

Lines including the word *issue* occur frequently in the category across the disciplinary/institutional sub-corpora because they are summarising a particular 'problematic' state of affairs. Similarly, words following **this** such as *fact* or *matter* usually fall into this grouping as summarising a 'situation ' or 'circumstance'.

Also included in this grouping are lines (occurring in the samples from the two 'History' sub-corpora illustrated in Table 134 below) containing words such as *period* or *phase* which do not denote a particular process but rather a context of time within which processes take place.

... **this** period of the Revolution ... [HIST-Y]
 ... **This** new phase ... [HIST-X]

Table 134: *this* + SITUATION OR CIRCUMSTANCE (2)

This is the most problematic ‘encapsulation’ grouping. More often than for other groupings an examination of co-text was needed to determine a line’s membership. An example is given in Table 135 below in which an examination of the preceding co-text shows that *centrifugal pattern* summarises the situation or state of affairs of the Islamic faith rather than encapsulating an ‘object’.

... it is apparent that "Islam" does not constitute a coherent monolithic threat fundamentally because "Islam" is a fragmented religion in itself. The multitude of Islamic movements, sects, belief systems that are prevalent within "Islam" is a reflection of **this** centrifugal pattern. ...

Table 135: *this* + SITUATION OR CIRCUMSTANCE (3)

4.7.3.6 ‘Encapsulation as Sense’

This small category contains sequences encapsulating a feeling, emotion or mental state (see Table 136 below). This category includes lines containing words such as ‘*view*’ and ‘*ideology*’ when they are denoting the perspective of an actor within the field or subject of study rather than a scholarly perspective on the field/subject.

... **this** desire to belong ...
 ... **this** predisposition ...
 ... **This** gendered fear ...
 ... **this** spirit ...
 ... **this** moral dilemma ...

Table 136: *this* + SENSE

4.7.4 'Commentary'

In the following groupings the semantic sequences including **this** (again, in most cases at the beginning of the sequence) do not *encapsulate* part or all of the preceding statement but instead make a *comment* about it. The keyword **this** on its own refers back to the content of the previous statement and is followed by a semantic element entailing a particular type of commentary on that content. These commentaries have been divided into seven semantic groupings explained below.

4.7.4.1 'Evaluation of preceding statement'

In the largest of these 'commentary' categories semantic sequences beginning with **this** perform the function of evaluating the content of the preceding statement. Phrases/clauses which perform this function take such a variety of lexicogrammatical shapes that they would create an unworkably large taxonomy of semantic sequences. Therefore, instead, the strings of words following **this** in all sequences in this group have been labelled with the 'catch all' semantic element 'EVALUATION' (the same has been done for all subsequent 'commentary' groupings except for the 'Exemplification' grouping).

The majority of 'evaluation' sequences contain a form of the verb 'to be' directly following **this** in most cases *is* or *was*. The majority of these sequences also contain evaluative adjectives (e.g. *quintessential*, *too much*, *important* highlighted in Figure 4.99 below) and in a few instances evaluative adverbs (e.g. *indeed* and *undeniably* highlighted in Table 137 below).

| |
|---|
| ... this <u>is</u> the quintessential nature of the African system ... |
| ... This <u>seems to give</u> too much credence to interpretation ... |
| ... This <u>was an</u> important validation of the principle of Kompetenz-Kompetenz ... |
| ... this <u>indeed</u> is the substance of the play ... |
| ... this <u>has</u> undeniably been its role since the beginning. ... |

Table 137: **this** + EVALUATION (1)

Three sequences included in this grouping in the pooled sample do not contain evaluative adjectives or adverbs (in Table 138 below) but their function is nonetheless obviously evaluative.

... this cannot count as empirical evidence ...
 ... this is the stripping down of history to its bare bones ...
 ... this seems to ignore that majority of developing world's population is concentrated in the key emerging markets countries...

Table 138: *this* + EVALUATION (2)

4.7.4.2 'Interpretation of preceding statement'

This grouping is differentiated from the 'Evaluation of preceding statement' grouping by the fact that phraseologies here are not giving a judgement of the preceding content, but instead explaining its meaning or implication (see Table 139 below). The word *why* occurs in a number of these lines, as do verbs with an interpreting function such as *demonstrate*, *disprove*, *mean*, *reflect*, *suggest* and *support*.

... This is perhaps why the creation of the hydrogen bomb seemed more questionable ...
 ... this demonstrates that Veblen is no disciple of classical economy ...
 ... This may reflect concern about the prosecution of dissenting states' citizenry ...
 ... This means the slave can create his identity through contradistinction ...
This would not be to suggest that international credibility was insignificant ...

Table 139: *this* + INTERPRETATION

A sub-category within this grouping, not large enough in either the pooled or disciplinary/institutional sub-corpora to be an independent grouping, shows how the idea contained within the statement preceding *this* relates to other ideas (illustrated in Table 140 below). This has been categorised for the purposes of this analysis as a form of 'interpretation'.

... this coheres with the elitism of the movement as a whole. ...
 ... this is more in line with modern, liberal conceptions of nationalism ...
 ... This taps into the recent trend in postmodernism ...
 ... this was of great contrast to Arnold's social thought ...

Table 140: *this* + INTERPRETATION (RELATIONSHIP TO OTHER IDEAS)

4.7.4.3 'Addition to preceding statement'

In this grouping illustrated in Table 141 below, the sequence beginning with ***this*** simply adds further information of a primarily factual nature to the content of the previous statement often in the form of more detail or context.

... this became a focusing objective for Muslim action and triumph ...
 ... This did not occur in Ghana ...
 ... this has been particularly popular within the UN framework ...
 ... this involves an interpersonal encounter ...
 ... This is also in the light of sovereignty becoming a widely important issue ...

Table 141: *this* + ADDITION

4.7.4.4 'Effect of preceding statement'

In this semantic category the sequence beginning with ***this*** explains the effect or impact of the content of the preceding statement (see Table 142 below).

... This created several negative effects of which the World Bank and the IMF struggled to justify and defend. ...
 ... this did not stop them from disobeying papal authority ...
 ... this in turn would highly increase the risk of, if not trigger, the 'slippery slope' into even more divisive and destabilising conflicts. ...
 ... this pleased the patient ...
 ... this has had an impact in reducing the worst cases of pollution across Europe ...

Table 142: *this* + EFFECT (1)

Some sequences within this grouping contain obvious 'effect' language such as *has resulted in*, *has had an impact in* and *led to*. One line contains the verb *'meant'*

but is grouped here rather than under ‘interpretation of preceding statement’ because in this instance it is judged to be denoting an effect rather than an interpretation (see Table 143 below).

... Although **this** usually meant shorter sentences in prison ...

Table 143: *this* + EFFECT (2)

4.7.4.5 ‘Exemplification of preceding statement’

In this grouping the sequence beginning with **this** provides an example of the content of the previous statement. There are two semantic sequences which realise this semantic motif. The first, ‘**this** + EXEMPLIFICATION’, includes the wide variety of ways an example is given by the string of words following **this**. The phrase *can be seen in* occurs in one third of the lines within this grouping in the pooled sample.

... **This** can be seen in the constant attempt by local elites to balance imperial friendly policy-making with domestic pressures ...
 ... **this** happened to some extent with the formation of the European Monetary Union ...
 ... **This** is very much the case in Pakistan ...

Table 144: *this* + EXEMPLIFICATION

The second involves a phrase including **this** which expresses the idea of exemplification followed by an example to illustrate the content of the statement **this** refers to.

... A good example of **this** was the 1882 suggestion, in Britain, of building a tunnel to connect Britain and France, The Channel Tunnel.
 ... an example of **this** is Leonardo of Bertipaglia ...
 ... a recent example of **this** being EU sanctions on imported clothes from China in order to protect European businesses ...

Table 145: EXEMPLIFICATION PHRASE [INCLUDING *this*] + EXAMPLE

4.7.4.6 'Cause of preceding statement'

In this grouping the sequence beginning with **this** provides a causal explanation for the content of the previous statement. Four of the seven sequences in this grouping include the word *because* as part of the 'CAUSE' element. Four of the seven lines in this grouping contain the word *because*. Two contain the phrasal verb *attributed to*.

| |
|---|
| ... This <u>cannot be solely attributed to the ICC prosecutions</u> ... |
| ... This <u>is because conspicuous consumption is believed to reflect the possession of socially superior and admirable traits</u> ... |
| ... this <u>is because of the significance of anarchy as a counter force</u> ... |
| ... this <u>probably had more to do with the nature of the challenges he faced than his own political preferences.</u> ... |
| ... this <u>was largely because of de Mountfort's experience and skill as a tactician.</u> ... |

Table 146: *this* + CAUSE

4.7.5 Semi-fixed expression - time reference

This grouping contains reasonably fixed phrases expressing the idea of a particular point in time in each case referring to the point in time expressed in the text preceding **this**.

| |
|---------------------------------|
| ... <u>At this junction</u> ... |
| ... <u>At this point</u> ... |
| ... <u>at this stage</u> ... |
| ... <u>at this time</u> ... |

Table 147: *at/to/till* + *this* + POINT-IN-TIME ENCAPSULATION

4.7.6 Semi-fixed expression – *in this manner/respect/sense/way*

Similarly, this grouping contains reasonably fixed phrases with the words *manner*, *sense*, *respect* and *way* encapsulating an idea expressed in the previous section of text.

| |
|--|
| ... <u>In</u> this <u>sense</u> ... |
| ... <u>in</u> this <u>way</u> ... |

Table 148: *in + this + MANNER*

4.7.7 Comparison/contrast across disciplinary and institutional sub-corpora

Table 149 below shows the frequencies of semantic groupings for **this** across the four disciplinary/intuitional sub-corpora. All potential frequency differences are differences between disciplines.

| Category | Pooled | | HIST -X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|---|------------|---------------|------------|---------------|------------|---------------|------------|---------------|------------|---------------|
| 1. 'Encapsulation as scholarly angle' | 32 | 10.67% | 28 | 9.33% | 23 | 7.67% | 34 | 11.33% | 40 | 13.33% |
| 2. 'Encapsulation as object' | 30 | 10.00% | 46 | 15.33% | 47 | 15.67% | 32 | 10.67% | 29 | 9.67% |
| 3. 'Encapsulation as process' | 29 | 9.67% | 39 | 13.00% | 35 | 11.67% | 36 | 12.00% | 17 | 5.67% |
| 4. 'Encapsulation of the essay text itself' | 20 | 6.67% | 5 | 1.67% | 13 | 4.33% | 28 | 9.33% | 32 | 10.67% |
| 5. 'Encapsulation as circumstance or state' | 16 | 5.33% | 9 | 3.00% | 20 | 6.67% | 21 | 7.00% | 24 | 8.00% |
| 6. 'Encapsulation as sense' | 10 | 3.33% | 10 | 3.33% | 14 | 4.67% | 6 | 2.00% | 5 | 1.67% |
| TOTAL - ENCAPSULATION | 137 | 45.67% | 137 | 45.67% | 152 | 50.68% | 157 | 52.33% | 147 | 49.01% |
| 8. 'Evaluation of preceding statement' | 39 | 13.00% | 22 | 7.33% | 22 | 7.33% | 7 | 2.33% | 24 | 8.00% |
| 9. 'Interpretation of preceding statement' | 22 | 7.33% | 23 | 7.66% | 21 | 7.00% | 44 | 14.66% | 38 | 12.66% |
| 10. 'Effect of preceding statement' | 17 | 5.67% | 17 | 5.67% | 16 | 5.33% | 15 | 5.00% | 14 | 4.67% |
| 11. 'Addition to preceding statement' | 16 | 5.33% | 21 | 7.00% | 21 | 7.00% | 15 | 5.00% | 11 | 3.67% |
| 12. 'Exemplification of preceding statement' | 9 | 3.00% | 3 | 1.00% | 4 | 1.33% | 14 | 4.67% | 11 | 3.67% |
| 13. 'Cause of preceding statement' | 7 | 2.33% | 6 | 2.00% | 5 | 1.67% | 0 | 0% | 3 | 1.00% |
| TOTAL - COMMENTARY | 114 | 38.00% | 95 | 31.66% | 91 | 30.33% | 97 | 32.33% | 106 | 35.33% |
| 14. Semi-fixed expression - 'Point in time' | 7 | 2.33% | 11 | 3.67% | 23 | 7.67% | 10 | 3.33% | 6 | 2.00% |
| 15. Semi-fixed expression – <i>in this manner/respect/sense/way</i> | 4 | 1.33% | 4 | 1.33% | 5 | 1.67% | 3 | 1.00% | 15 | 5.00% |
| 16. Other | 42 | 14.00% | 56 | 18.67% | 32 | 10.67% | 35 | 11.67% | 31 | 10.33% |

Table 149: Frequencies of *this* phraseologies across disciplinary/institutional sub-corpora

4.7.7.1 Similarities

A strong continuity across all four sub-corpora is to be found in the ratio of 'encapsulation' to 'commentary' phraseologies. As can be seen above in Table 4.21 the 'encapsulation' category is always noticeably larger, making up between 45.67% and 52.33% of each sub-corpora random sample. The 'commentary' category makes up 31.66% to 38% of the sub-corpora random samples.

4.7.7.2 Variations

A number of disciplinary variations emerged in the first random samples for each of the disciplinary/institutional sub-corpora.

The frequency of 'Encapsulation as scholarly angle' is slightly higher in the POL/IR sub-corpora than the HIST sub-corpora. There are more than double the instances of 'Encapsulation of the essay text itself' within the POL/IR sub-corpora. Frequency of 'Meaning of preceding statement' lines and 'Exemplification of preceding statement' lines are also noticeably higher in POL/IR although the latter is a small group in all four sub-corpora. The frequency of 'Meaning' lines is dramatically higher in POL/IR than HIST.

In the HIST sub-corpora the 'Encapsulation of object' grouping is approximately a third larger than in it is in the POL/IR sub-corpora and there are also noticeably more 'Addition to previous statement' lines. There are also more 'cause of preceding statement' lines although this is an extremely small grouping across all four sub-corpora.

Second random samples for each disciplinary/institutional sub-corpora were analysed and results can be seen in Tables 150 to 4.28 below.

| 'Encapsulation as scholarly angle' | HIST -X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|------------------------------------|---------|-------|--------|-------|----------|--------|----------|--------|
| Sample 1 | 28 | 9.33% | 23 | 7.67% | 34 | 11.33% | 40 | 13.33% |
| Sample 2 | 19 | 6.33% | 16 | 5.33% | 30 | 10.00% | 27 | 9.00% |

Table 150: Comparison of first and second sample frequencies for 'Encapsulation as scholarly angle'

| 'Encapsulation as object' | HIST -X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|---------------------------|---------|--------|--------|--------|----------|--------|----------|--------|
| Sample 1 | 46 | 15.33% | 47 | 15.67% | 32 | 10.67% | 30 | 10.00% |
| Sample 2 | 39 | 13.00% | 40 | 13.33% | 28 | 9.33% | 30 | 10.00% |

Table 151: Comparison of first and second sample frequencies for 'Encapsulation as object'

| 'Encapsulation as the essay text' | HIST -X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|-----------------------------------|---------|-------|--------|-------|----------|-------|----------|--------|
| Sample 1 | 5 | 1.67% | 13 | 4.33% | 28 | 9.33% | 32 | 10.67% |
| Sample 2 | 10 | 3.33% | 12 | 4.00% | 23 | 7.67% | 30 | 10.00% |

Table 152: Comparison of first and second sample frequencies for 'Encapsulation as the essay text'

| 'Interpretation of preceding statement' | HIST -X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|---|---------|-------|--------|-------|----------|--------|----------|--------|
| Sample 1 | 23 | 7.66% | 21 | 7.00% | 44 | 14.66% | 38 | 12.66% |
| Sample 2 | 29 | 9.66% | 27 | 9.00% | 41 | 13.66% | 38 | 12.66% |

Table 153: Comparison of first and second sample frequencies for 'Meaning of preceding statement'

| 'Addition to preceding statement' | HIST -X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|-----------------------------------|---------|-------|--------|-------|----------|-------|----------|-------|
| Sample 1 | 21 | 7.00% | 21 | 7.00% | 15 | 5.00% | 11 | 3.67% |
| Sample 2 | 21 | 7.00% | 16 | 5.33% | 14 | 4.67% | 13 | 4.33% |

Table 154: Comparison of first and second sample frequencies for 'Addition to preceding statement'

| 'Exemplification of preceding statement' | HIST -X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|--|---------|-------|--------|-------|----------|-------|----------|-------|
| Sample 1 | 3 | 1.00% | 4 | 1.33% | 14 | 4.67% | 11 | 3.67% |
| Sample 2 | 4 | 1.33% | 7 | 2.33% | 9 | 3.00% | 13 | 4.33% |

Table 155: Comparison of first and second sample frequencies for 'Exemplification of preceding statement'

| 'Cause of preceding statement' | HIST -X | | HIST-Y | | POL/IR-X | | POL/IR-Y | |
|--------------------------------|---------|-------|--------|-------|----------|-------|----------|-------|
| Sample 1 | 6 | 2.00% | 5 | 1.67% | 0 | 0% | 3 | 1.00% |
| Sample 2 | 7 | 2.33% | 5 | 1.67% | 2 | 0.67% | 2 | 0.67% |

Table 156: Comparison of first and second sample frequencies for 'Cause of preceding statement'

As can be seen from these tables, in each case the second samples confirmed the presence of disciplinary differences identified in the first samples.

4.8 Conclusion

This chapter has described in detail the results of analysis of the semantic patterning of five grammatical words which are statistically key across the four disciplinary/institutional sub-corpora, *of*, *and*, *that*, *as* and *this*. In Chapter 5 the implications of these findings will be explored. This will include discussion of what insights the findings of similarity and difference in the current study may provide with regards to the nature of student writing within the two target disciplines and also within the wider category of student Humanities and Social Sciences writing. It will also include discussion of the implications for EAP knowledge and pedagogy and of the strengths and limitations of the methodology as it has been adapted and used within the current project.

Before turning to this discussion in Chapter 5, however, it is worth noting a small observation about the analytical procedure here. In the qualitative process of semantically categorising the phraseological patterning of target keywords, there will inevitably be lines in the concordance sample that cannot be satisfyingly

grouped and are therefore relegated to 'other'. Just above the level of 'other' there will also usually be a grouping that the researcher is left feeling a little uneasy about. This uneasiness comes from the fact this grouping is slightly less robust, slightly more ambiguous and 'catch all' in nature than other groupings for the keyword. The lines within it have a strong *enough* connection that an argument can be made to semantically group them rather than discard them. However, the semantic function of the grouping is broader and more abstract than others so possibly tells us less that is useful about the target discourse, and the connection between members of the grouping could also be criticised for being slightly tenuous. Within the current study examples of such a category would be 'listed miscellany' in **and**, PHENOMENON + **that** + DETAIL in **that** and 'encapsulation as object' in **this**. Rather than a 'limitation', I would argue this is an inevitable end point of this analytical process and as such a useful indicator for a researcher using this method of where to draw the lines in terms of level of classificatory delicacy.

Chapter 5: Discussion

This chapter starts with a reminder of the research questions with the final answers to these provided in sections 5.2 and 5.3. Section 5.2 outlines and explains both continuities across the four sub-corpora and disciplinary variations between them, the former with reference to features shared by disciplines which fall within a 'Humanities and Social Sciences' grouping (Durrant, 2015), the latter with reference to particular preferences and foci within either History or Politics/International Relations. Section 5.3 provides a summative discussion of the relative influence of discipline and institution on both continuity and variation. Section 5.4 explores potential contributions and implications of the research with regards to 'coverage' in terms of the proportion of text accounted for by the identified phraseologies, the role and nature of disciplinary language and related pedagogical implications, and what light the findings shed on the EGAP (English for General Academic Purposes) and ESAP (English for Specific Academic Purposes) debate. Finally section 5.5 reflects on the strengths and limitations of the methodology and of the research design as a whole.

5.1 Aims and Research Questions

The aim of this project was to investigate whether disciplinary differences could be identified at the lexico-grammatical level between good third-year undergraduate writing in two relatively close 'soft' disciplines, History and Politics/International Relations. To account for the possibility that factors other than discipline may have a strong influence on the nature of student writing at this level, a cross-disciplinary, cross institutional research design was adopted which involved investigating the two target disciplines at two institutions.

Four sub-corpora were created each containing good third-year texts from the Essay genre family (Nesi and Gardner, 2012) drawn from one of the two disciplines at one of the two institutions. A bottom-up corpus-driven approach was taken in

which a slightly adapted version Groom's (2007) methodology was used involving the semantic analysis of the phraseologies occurring with statistically key grammatical words. Five grammatical words key across all four sub-corpora, ***of, that, as, this*** and ***and*** were analysed in this way to answer the following research questions.

1a. In a comparison/contrast between successful third-year undergraduate texts written within the History degree programme and those written within Politics/International Relations degree programme at Institution X, what *similarities* can be identified in the lexico-grammatical and semantic patterns occurring around positively key closed-class grammatical items?

1b. In a comparison/contrast between successful third-year undergraduate texts written within the History degree programme and those written within Politics/International Relations degree programme at Institution X, what *differences* can be identified in the lexico-grammatical and semantic patterns occurring around positively key closed-class grammatical items?

2. To what extent are similarities and differences identified between writing in programmes at institution X replicated in a comparison/contrast of successful third-year writing between the equivalent degree programmes within Institution Y?

3. What arguments can be made from the findings of the first two questions in terms of the extent to which either 'discipline' and/or local institutional context seem to have a stronger relationship to any similarities or differences found among the lexico-grammatical patterns/phraseology exhibited by the different groups?

Questions 1(a), 1(b) and 2 are three components of a broader question about patterns of continuity and difference across the four disciplinary/institutional sub-corpora. Question 3 asks how the answers to 1(a), 1(b) and 2 can be interpreted. In section 5.2, the findings of continuity and variation across the disciplinary/institutional sub-corpora described in Chapter 4 are summarised before

an attempt is made to explain them. Continuities will be addressed first in section 5.2.1 then variations will be addressed in section 5.2.2.

5.2 Similarities and variations across the four disciplinary/institutional sub-corpora

This section will firstly outline and attempt to account for phraseological continuities across the four disciplinary/institutional sub-corpora. It will then outline and attempt to account for variations across the four sub-corpora.

5.2.1 Similarities

Broad patterns of similarity across both discipline and institution make up a greater proportion of the overall results than do patterns of difference. The largest of these continuities are summarised in Table 157 below.

| | |
|--------------------|--|
| <i>of</i> | <ul style="list-style-type: none"> • > 90% of phraseologies realised grammatically through noun/noun phrase + <i>of</i> + noun/noun phrase nominalisation • Similar patterns of frequency with the eight largest groupings occurring in all sub-corpora with a frequency of at least 2%: PROPERTY + <i>of</i> + PHENOMENON; PROCESS + <i>of</i> + OBJECT; PROCESS + <i>of</i> + ACTOR; PROPORTION OR AMOUNT + <i>of</i> + PHENOMENON; CONCEPTUALISATION + <i>of</i> + PHENOMENON; QUALITY + <i>of</i> + PHENOMENON; PHENOMENON + <i>of</i> + INSTANCE OR NATURE; CONCEPTUAL LABELLING + <i>of</i> + CONCEPT |
| <i>and</i> | <ul style="list-style-type: none"> • Between 78% and 87% of phraseologies consist of ‘listed phenomena’ • Similar patterns of frequency with seven groupings occurring in all sub-corpora with a frequency of at least 2%: PROCESS + <i>and</i> + PROCESS; ABSTRACT CONCEPT + <i>and</i> + ABSTRACT CONCEPT; QUALITY + <i>and</i> + QUALITY; CLASSIFIER + <i>and</i> + CLASSIFIER; ‘listed people’; LOCATION + <i>and</i> + LOCATION; ‘addition’ |
| <i>that</i> | <ul style="list-style-type: none"> • The semantic motif ‘voice’ constituting between 50% and 60% of all phraseologies: <ul style="list-style-type: none"> ○ ‘the voice of the essay writer’ (25% to 39%) ○ ‘the voice of others’ (18% to 29%) • The sequence PHENOMENON + <i>that</i> + DETAIL constituting approximately 20% of all phraseologies • Phraseologies with <i>the fact that</i> constituting between 4% and 6% of all phraseologies |
| <i>as</i> | <ul style="list-style-type: none"> • The semantic motif ‘interpretation’ constituting the largest or second largest grouping (between 15% and 27%) & the semantic motif ‘interpretation: perception’ constituting the largest sub-category of ‘interpretation’ (between 9% and 17%) • All other semantic motifs occurring with at least 2% frequency across all sub-corpora: ‘exemplification’; ‘addition’; ‘attribution’; ‘function’; ‘role’; ‘conceptualisation’; ‘similarity’; ‘extent’; ‘simultaneity’ |
| <i>this</i> | <ul style="list-style-type: none"> • Similar relative proportions of ‘encapsulation’ and ‘commentary’: • The umbrella semantic motif ‘encapsulation’ constituting 45-52% of all phraseologies. Presence across all sub-corpora of ‘encapsulation’ of: scholarly angle; object; process; the essay text itself; situation or circumstance; sense • The umbrella semantic motif ‘commentary’ constituting 31-38% of all phraseologies. Presence across all sub-corpora of the following types of commentary: evaluation; interpretation; addition; describing effect; exemplification; attributing cause |

Table 157: Patterns of continuity in the semantic motifs or semantics sequences of *of*, *and*, *that*, *as* and *this*

Broadly speaking, the greater proportion of findings within the current study thus point not to disciplinary or institutional difference, but, instead, to commonalities shared across two disciplines. Student essay writers in History and student essay writers in Politics/International Relations at both institutions seem to be ‘doing’ similar things semantically within their texts. This finding coheres with the similarities between History and Politics found by Durrant (2015) in his study of 4-grams and lexical bundles within BAWE texts: both disciplines were clustered in a

‘Humanities and Social Sciences’ grouping and within that grouping they formed a further clustered sub-grouping of Law, History and Politics (2015, p. 11) indicating a relatively close relationship even within this disciplinary cluster.

It should not be automatically concluded, however, that *all* identified phraseological continuities between History and Politics/International Relations are directly attributable to their ‘Humanities and Social Sciences’ character. For this to be true all their common phraseological features would have to be found either exclusively or always in noticeably higher numbers within ‘Humanities and Social Sciences’ texts.

This is unlikely to be the case; for example, in Hardy and Römer’s (2013) Multi-Dimensional (MD) analysis of the lexico-grammatical features of writing across disciplines in the MICUSP corpus discussed in detail in Chapter 2, whilst many ‘dimensions’ had positive or negative associations with either ‘soft’ fields collectively (Humanities and Arts, Social Sciences) or ‘hard’ fields collectively (Physical Sciences, Biological and Health Sciences), there were also overlaps, and differences between ‘soft’ and ‘hard’ subjects tended to be a matter of degree better expressed along a cline than in terms of absolute differentiated categories.

The current study only involves analysis of essays in the two ‘soft’ disciplines of History and Politics/International Relations. Consequently there are no internal means of comparing and contrasting the phraseological features in these ‘soft’ subjects with the phraseological features of student essay writing in ‘hard’ and/or scientific fields. For this reason an external source of information is needed in order to have a set of criteria by which to judge what particular phraseologies common to History and Politics/International Relations might be argued to be especially or specifically attributable to their status as part of ‘Humanities and Social Sciences’.

The most obvious choice for this external point of reference for this purpose is Durrant’s (2015) conclusions regarding the features of the ‘Humanities and Social

Sciences’ grouping in contrast to the ‘Science and Technology’ grouping which emerge from his analysis of N-grams and lexical bundles in the BAWE corpus (see Table 158 below). Durrant’s conclusions relate to student texts, and so are a better reference for the present study than would be features of the ‘soft’ disciplines as described by, say, Becher and Trowler (2001) or Hyland (e.g. 2009) whose main focus is professional academic texts. Durrant’s conclusions also have the advantage of emerging from a bottom-up analysis of the BAWE data representing a grounded source of descriptive features for specifically student writing in the ‘soft’ disciplines.

- | |
|--|
| <ul style="list-style-type: none"> • a focus on abstract constructs • a focus on historical moments/points in process • emphasizing the role of unique autonomous agents in processes that are difficult to control • showing multiple contingent viewpoints • evaluation: expressing attitudes towards objects of study • establishing centrality • setting things in interpretive/limiting contexts • setting ideas in relationships with each other |
|--|

Table 158: Common features of student ‘Humanities and Social Sciences’ texts (Durrant (2015))

Where and how the semantic sequences and semantic motifs identified in the current study map or link to Durrant’s identified features of student Humanities and Social Sciences writing is discussed below. All points listed in Table 158 are discussed with the exception of ‘establishing centrality’, a feature which did not appear to any noticeable extent within the phraseologies that emerged in the current study.

5.2.1.1 A focus on abstract constructs

Durrant identifies ‘a focus on the physical world vs. a focus on abstract constructs’ as being a ‘primary contrast’ between student Science and Technology writing on the one hand and student Humanities and Social Sciences writing on the other (2015, p. 21). A focus on abstract constructs is reflected very strongly within this study’s phraseological findings for the two Humanities and Social Sciences disciplines of History and Politics/International Relations.

In exploring the extent of focus on abstract constructs in the current study, an important starting point is a clear delineation of what is and what is not to be considered 'abstract'. Martin's (1997, p. 30) network of Thing types differentiates metaphorical and abstract from concrete nouns in a rich taxonomy from 'concrete specific' to 'metaphoric quality'. However, this level of delicacy and complexity is unnecessary and, arguably, would be counterproductive for the current purposes, so a more straightforward approach will be taken here. The adjectival form of 'abstract' is defined by the Oxford English Dictionary as '[e]xisting in thought or as an idea but not having a physical or concrete existence' while an *abstract noun* is defined as 'denoting an idea, quality, or state rather than a concrete object' (as mentioned already in section 4.4.2.2 of Chapter 4).

In the phraseologies of the closed class keywords which are the focus of this study words such as *culture*, *liberty* and *legacy* fall definitively within what is abstract, and thus represent 'abstract constructs', as do noun phrases such as *artistic style*, *leftist milieu* and *geopolitical tensions*, whilst, in contrast, *a green toad*, *water* and *Willy Brandt* are all entities with a concrete material reality and thus non-abstract.

Processes or circumstances expressed as ideas, such as *female medical aid*, *the New York protests* and *the challenges he faced* can also be considered abstract constructs. Some nouns and noun phrases, however, arguably fall somewhere on a cline between abstract and non-abstract: *the international criminal court* could be considered as a concrete configuration of specific human actors and locations or as a more abstract international legal 'mechanism', whereas the phrases *geopolitical structures*, *political organisations* and *orange-type regimes* fall more firmly in the 'abstract' category as construing conceptually rather than materially the mechanisms of political control; *the elite* or *the least advantaged* could, depending on context, represent concrete groups within a specific wider population or more abstractly represent concepts which can be applied within an explanation of any societal context.

For the purposes of this analysis, any element within an identified instance of a semantic sequence which can be argued, within the context it appears, to construe *particular material* entities, such a human individual or group, a geographical location, or object/set of objects will be considered to be non-abstract.

Whilst a focus on abstract constructs is seen in many of the phraseologies across all grammatical keywords analysed within this study, it is especially predominant in **of** where there is a strong focus on the abstract across all phraseological categories. The names of many of the semantic sequences for **of** automatically prescribe an abstract focus: **PROCESS + of + OBJECT** and **PROCESS + of + ACTOR** prescribe that the elements preceding **of** are abstracted verbal processes; **CONCEPTUALISATION + of + PHENOMENON** and **EVALUATION + of + PHENOMENON** also automatically render the elements preceding **of** as abstract in focus, and the naming of the **CONCEPTUAL LABELLING + of + CONCEPT** semantic sequence denotes an entirely abstract focus.

When we remember that the word **of** on its own constitutes approximately 4-5% of the content of student essay texts across disciplines and institutions within the study, we can start to get a sense of the reach of this attention to the abstract within student writing in both History and Politics/International Relations.

The level of focus on the abstract within phraseologies grouped under the semantic sequence label **PROPERTY + of + PHENOMENON** (the largest grouping in the pooled sample) is representative of the dominance of this focus on abstract constructs across all semantic sequences for **of**. For illustrative purposes, it is thus worth carrying out a detailed examination of the semantic sequences within this grouping. The fifty five lines grouped as **PROPERTY + of + PHENOMENON** sequences in the pooled sample are reproduced in Table 159 below.

the **appendages** of Bourbon monarchy
 the contextual background of each ethnic conflict
 the bounty of **China**
 the culture of **the artisan**
 the productive forces of **the country**
 the **fruits** of the State'
 the future of economic progress
 the social and cultural **landscape** of the ancien régime
 the lexicon of feminist struggle
 the political organisations of **the continent**
 the riches of the State
 the rights of **the individual**
 the multiple skills of **a single man**
 the uneasy theme and structure of the Brutus
 the geopolitical structures of **the world**
 the overriding geopolitical tensions of the early Cold War
 the themes of this essay
 new topics of the international agenda
 the requirements and logic of the hour.'
 the personality of **Willy Brandt**
 the personality of **the Statesmen**.
 the role of institutions
 the future workings of **the Council**.
 the greatest achievement of **the international criminal court**
 a manifesto of the revolution in Russia
the foot of **a green toad**
 the **surface** of **water**,
 occupation of **the criminal vagrant**
 the artistic style of Ranke's history
 an internationalised **theatre** of the Cold War conflict
 the basis of a conflict with the "West".
 the position of **the least advantaged**
 the rites of religion'
 the **weapons** of democracy
 the source of wealth.
 the discourses of the New York protests
 the discourse of OWS
 The leftist milieu of the early Nasser regime
 "the qualitative composition" of the proletarian component.
the outskirts of **the great city**
 the very top of **the chain**.
 mere components of the homogenising discourse.
 the imperialist aspiration of **Germany**
 the attitudes of **the elite**
 the dreamtime beliefs of **the Aborigines**
 the nature of female medical aid
 the nature of the challenges he faced
 the nature of relations between the two
 the entire nature of the post-war international settlement
 the nature of **the metropolis**
 the most important part of his legacy
members of **the great families**.
 the liberty of everybody
 the predictable crisis of the current orange-type regimes
 the details of events

Table 159: PROPERTY + *of* + PHENOMENON sequences from the pooled sample

A quick scan of Table 159 is likely to give the reader an immediate impression of the predominance of abstract concepts populating both the **PROPERTY** element preceding *of* and the **PHENOMENON** element following *of* in these sequences.

When the **PROPERTY** element is examined more closely, only nine of the fifty-five lines contain potentially concrete nouns (highlighted in Table 159 above). However,

appendages, fruits, landscape, theatre, and weapons are all used in a metaphorical sense which renders them abstract, leaving only *foot, surface, outskirts* and *members* which can be argued to represent tangible material entities (N.B. *bounty* and *riches* have been judged to be abstract due to their emphasis not on particular material possessions but on relative degrees of conceptualised *wealth*).

When the **PHENOMENON** element is examined closely, we find three potential representations of material objects, *toad, water* and *chain*, but the last is used in a metaphorical and thus abstract sense. There are seven specific geographical/spatial locations – *China, the country, the continent, the world, the great city, Germany* and *the metropolis* (the last referring specifically to London). Two elements have been deemed to simultaneously represent specific locations and specific groups of people, *the Council* and *the international criminal court*. There are ten representations of human beings either individually or as a group – *the artisan, the individual, a single man, Willy Brandt, the statesmen, the criminal vagrant, the least advantaged, the Aborigines, the elite* and *the great families*. However, on a closer examination of these, we find that *the artisan, the individual, a single man, the criminal vagrant* and *the least advantaged* are all used in an abstract sense to express a category or model of human being rather than specific actual human entities (in contrast *the elites*, for example, are here referring to a portion of a specific population, that of Great Britain, in a specific time period).

Thus, out of 110 possible semantic element positions in the fifty-five semantic sequences in this grouping in which a semantic element could represent a material entity, only twenty elements (18%) can be very confidently said to do so (these are underlined as well as highlighted in Table 5.3). This leaves the overwhelming majority of semantic elements within the **PROPERTY + *of* + PHENOMENON** grouping, 82% in total, representing abstract constructs.

This pattern of representation of predominantly abstract constructs is repeated across all semantic sequence categories for *of* in the pooled sample. In fact, only five sequences, 2% of all categorised phraseologies for *of*, consist of entirely

concrete elements, so few that they can be listed here: the foot *of* a green toad, the surface *of* water, the outskirts *of* the great city, David's sketch *of* Marie Antoinette, and the Straits *of* Gibraltar.

This focus on abstract constructs can also be identified within the phraseological analysis of *that* particularly within the PHENOMENON + *that* + DETAIL semantic sequence grouping which constitutes 70 of the 300 *that* concordance lines in the pooled random sample. A quick count finds six human (*Muslim figures, other leaders, those men, the number of men, many others, others*), one geographical (*a country*) and one material element (*smoke*) in the PHENOMENON position which can be considered non-abstract. Thus, fewer than 12% of the lines in this grouping have a non-abstract focus meaning 88% have an abstract focus.

A similarly high level of abstract and, indeed, metaphorical content can be seen within the phraseologies for *as* particularly in the semantic motifs 'function', 'role' and 'conceptualisation' illustrated in lines [1] to [4] below. In lines [1], [2] and [3] we see the concrete entities *individual, the Mediterranean, and Germany* construed metaphorically as a *building block, a gateway* and a *competitor* respectively, and in line [4] we see *power* anthropomorphised as an actor who has an *adversary*.

[1] ... the unified individual *as* key 'building-bloc' of any social theory of agency and culture ...

[2] ... *as* the "gateway to the Middle East", the Mediterranean ...

[3] ... Germany *as* an imperialist competitor ...

[4] ... that which power targets *as* its adversary ...

In the 'encapsulation' groupings for *this* a strong focus on abstract constructs is found again. Of the six 'encapsulation' groupings, five automatically denote a focus on abstract constructs: 'encapsulation *as scholarly angle*', 'encapsulation *as process*', 'encapsulation *as the essay text itself*' (an abstract discoursal object), 'encapsulation *as situation or circumstance*' and 'encapsulation *as sense*'. Only one

grouping, ‘encapsulation as object’, has the potential to contain phraseologies with a specific material focus. When this grouping in the pooled sample is examined, seven of the thirty lines contain representations of concrete material entities: *British outpost, extraordinary woman, grave, healer, painting, region* and *text*. Thus, only 5% of the ‘encapsulation’ lines as a whole contain non-abstract foci while in 95% the focus is abstract. The focus on abstract concepts in second largest semantic grouping for **and**, which accounts for between 7% and 13% of phraseologies in each of the sub-corpora, is obvious from its label, **ABSTRACT CONCEPT + and + ABSTRACT CONCEPT**.

5.2.1.2 A focus on historical moments/points in a process

Durrant finds that, in contrast to the predominant focus on physical location in student Sciences and Technology writing, student writing in the Humanities and Social Sciences locates events temporally within historical time or within a process (2015, p. 21). In the current study, this focus on temporal, particularly historical, processes can be found within particular semantic sequences and semantic motifs within the phraseologies for **of**, **as**, and **this**.

An emphasis on historical processes is found in many of the sequences within two of the largest groupings for **of**, **PROCESS + of + OBJECT** (illustrated in lines [5] and [6] below) and **PROCESS + of + ACTOR** (illustrated in lines [7] and [8]).

[5] ... creation **of** institutions at regional level ...

[6] ... the lived experience **of** the early revolution ...

[7] ... a complete breakdown **of** effective government ...

[8] ... The rise **of** nationalist identity in ethno-national conflicts ...

Within phraseologies for **as**, construal of historical processes in temporal relation to each other is found in the semantic motif ‘simultaneity’ (illustrated in lines [9] and [10] below).

[9] ... Czech cultural integrity faded **as** the fog of war descended over Europe ...

[10] ... there was a reappraisal of Ranke’s role in American historiography **as** the German interpretation crossed the Atlantic ...

Within **this**, emphasis on historical processes and experiences is found in the encapsulation of process’ category (illustrated in lines [11] to [13] and emphasis on historical context is found in the ‘encapsulation of situation or circumstance’ category (illustrated in lines [14] to [16]).

[11] ... **this** conflict ...

[12] ... **this** destructive drive ...

[13] ... **this** negative experience ...

[14] ... **this** imbalance ...

[15] ... **this** background of societal change ...

[16] ... **this** context ...

5.2.1.3 Emphasizing the role of unique autonomous agents in processes that are difficult to control

Durrant finds that, in contrast to the achieving of ‘tight control’ over foci within the student Science and Technology writing, student writing in the Humanities and Social Sciences involves ‘less determinative situations where influence is sought but not necessarily achieved’ (2015, p. 22). He contrasts the use of the bundle *the use of a* in Science and Technology (to refer to ‘any individual case of a generic set’) with the bundle *the use of the* in Humanities and Social Sciences (to refer to ‘specific

cases') to underline the contrast between 'Science and Technology writers' focus on interchangeable objects as instruments and on achieving tight control and Humanities and Social Sciences writers' focus on specific, autonomous agents and complex, uncontrollable situations' (ibid.).

In the current study this Humanities and Social Sciences emphasis on 'unique autonomous agents' within difficult-to-control processes is found in phraseologies within the 'voice of others' semantic motif for **that** and within the 'Interpretation' semantic motif for **as**.

In the 'voice of others' category for **that** a noticeable proportion of phraseologies, just under 20%, involve 'voices from the field' (as illustrated by lines [17] to [19] below)

[17] ... Fierlinger reported to the exile government on 4 January 1942 **that** 'Some pseudo-liberal regime will not ...

[18] ... In 1987 Gorbachev said **that** 'the time is ripe for abandoning ...

[19] ... a "religious zealot" who fervently opposed **that** western egalitarianism was unsuitable for Pakistan ...

These people arguably are unique autonomous agents within the difficult-to-control processes of History and Politics/International Relations, and, where they stand in relation to these processes and how they perceive these processes appear to be important foci for student writers in these two disciplines.

Within the 'listed people' semantic grouping for **and**, which constitutes between 6% and 10% of phraseologies across the four disciplinary/institutional sub-corpora, the majority of sequences involve unique autonomous agents from the field in focus (illustrated in lines [20] to [22]).

[20] ... doctors, apothecaries **and** grocers ...

[21] ... Bohr **and** Heisenberg ...

[22]... Mrs. Montague **and** the adherents of the Burlington School of Taste ...

The views of actors from the field, unique autonomous agents, also form a large proportion of the focus of ‘interpretation’ phraseologies for **as** (illustrated in [23] to [25] below). Again, we see these ‘agents’ acting in and reacting to national and international political and historical processes.

[23] ... interpreted by the Chinese **as** fervent American imperialism ...

[24] ... They saw neurasthenia **as** a chance to bargain with traditional patriarchy ...

[25]... Promoted by Jerry Rawlings, head of the Provincial National Defence Council, **as** an incredible opportunity for economic redevelopment ...

5.2.1.4 Showing multiple contingent viewpoints & setting ideas in relationship with each other

This section addresses two linked features which Durrant identifies in his study as distinguishing student writing in Humanities and Social Sciences from student writing in Science and Technology. Firstly, Durrant finds that student Humanities and Social Sciences writing distinctly involves ‘a discourse of contingent points of view and arguments which contrast with the straightforward reporting of ‘findings’ prevalent in Science and Technology’; in student Humanities and Social Sciences writing there is ‘emphasis ... on multiple interpretations and attributed statements’ (2015, p. 23). Relatedly, he finds that, in contrast to Science and Technology writers who ‘explain their case by building up chains of positive reasoning’ student Humanities and Social Sciences writers ‘clarify their argument by acknowledging and distinguishing their position from alternative inferences’ (ibid., p. 25).

In the current study these intertwined features of student Humanities and Social Sciences, those of both expressing as well as showing relationships between multiple viewpoints, are strongly reflected in a large range of the phraseologies.

Emphasis on multiple interpretations is found for **of** within the semantic sequences in the **CONCEPTUAL LABELLING + of + CONCEPT** category. In four of the twelve lines in this category within the pooled sample the **CONCEPTUAL LABELLING** element is plural (illustrated in lines [26] and [27]) thereby arguably carrying within it a sense of *more than one* viewpoint.

[26] ... notionss of identities ...

[27]... theoriess of original accumulation ...

Other sequences within this category can also be argued to carry an intrinsic sense of competing viewpoints (illustrated in lines [28] to [30]); *Foucault's*, as opposed to *anyone else's*, idea of resistance, *religious*, as opposed to *secular*, ideas about women, *a new* way of understanding power, as opposed to *previous* approaches.

[28] ... Foucault's conception of resistance ...

[29]... religious conceptions of women ...

[30]... a new understanding of power ...

The primary job of **that** within the student essay texts in this study is to express viewpoints: more than fifty percent of phraseologies perform a 'voice' function (illustrated in lines [31] to [35] below).

[31] ... Foucault stresses **that** discourse is associated with relations of power ...

[32]... Scholars of this theory believe **that** once cooperation between states is ...

[33] ... David's hypothesis **that** the Ranters did not exist ...

[34]... an interesting suggestion **that** states are becoming more uniform ...

[35]... I believe **that** the poor conditions of the workers ...

In concurrence with Durrant's findings, the phraseologies in this study are characterised by a predominance of 'contingent' language for reporting viewpoints

in that the language used acknowledges the possibility of competing/conflicting ideas. There are far more instances of reporting language which carries the implication of debate and the contested nature of knowledge than there are instances of reporting language which presents propositions as ‘fact’ such as the reporting verbs *show* and *find*.

A survey of phraseologies in the ‘voice of others’ category within the pooled sample yields only two examples of use of the verb *show* but multiple instances of forms of the verbs *believe*, *argue*, *claim* and *suggest*. Similarly, there are no instances of the use of *finding* but multiple instances of *argument* and *idea* along with range of other nouns which frame propositions as one among competing perspectives: *assumption*, *charge*, *hypothesis*, *interpretation*, *notion*, *outlook*, *perspective* and *suggestion*.

The expression of competing scholarly views of the field in focus can also be found in many of the lines in the ‘interpretation’ grouping for **as**, particularly those from the sub-grouping ‘interpretation: portrayal’ as seen in lines [36] and [37] below. It can also be seen in the semantic motif of ‘attribution’ for **as**, involving the attributing of a proposition to a particular author, shown in lines [38] and [39].

[36] ... Orthodox historians, such as Michael Schaller, strongly argue that, as of 1979, Sino-American action cannot be described **as** a Cold War ...

[37] ... They chart its inception **as** the realisation of “enlightened” principles espoused by the eighteenth-century philosophes ...

[38] ... but **as** Peter Reil has argued, his spontaneity is distinct from the mechanistic notions one associates with sciences. ...

[39] ... **As** English has commented, ‘It would be wrong to see 1981 as ...

The Humanities and Social Sciences characteristic of representing multiple views set in relationship to each other is also reflected in the semantic motif ‘encapsulation of

scholarly angle' for **this**, which represents between 7% and 14% of use of **this** across disciplinary/institutional sub-corpora (illustrated in lines [40] to [42]).

[40] ... **this** abstract argument ...

[41] ... **this** notion ...

[42] ... **this** thesis ...

As with the language used in 'voice of others' phraseologies for **that**, the language used in the majority of **CONCEPTUAL LABELLING** elements within semantic sequences in this category carries the implication of the existence alternative viewpoints: *argument, conception, interpretation, notion, opinion, theory, thesis* and *view*.

5.2.1.5 Evaluation

Durrant finds that student Humanities and Social Sciences writing has an 'inherently evaluative nature' a distinguishing feature of it being the importance of 'expressing attitudes towards its objects of study (2015, p. 23). In the current study this evaluative feature is found in particular semantic sequences and motifs within the phraseologies for **of**, **and**, **as** and **this**.

An explicit evaluative function is performed by the **QUALITY** element of the sequence **QUALITY + of + PHENOMENON** within the semantic categories for **of** as illustrated by lines [46] to [48] with the relevant language highlighted.

[46] ... the **advantages** of progress ...

[47] ... The **contradictory** nature of Britain's foreign policy ...

[48] ... the **effectiveness** of the Soviet as a body for socialist representation ...

An explicit evaluative function also is performed by both **QUALITY** elements the third largest semantic grouping for **and**, the sequence **QUALITY + and + QUALITY** (illustrated in lines [43] to [45] below).

[43] ... corruption and predatoriness ...

[44] ... appropriateness, efficiency and effectiveness ...

[45] ... jingoism and exclusivity ...

Evaluative language is also found within the element following **as** in many 'interpretation' phraseologies, especially 'interpretation: judgement', illustrated in lines [49] to [51] below.

[49] ... what was widely regarded as inappropriate international policy ...

[50] ... referring to the witch hunts as drastic examples of midlead [sic] gossip and rumours ...

[51] ... reviled by early Christian historians as 'the very incarnation of evil' ...

The function of evaluation is also central in phraseologies grouped within the semantic motif 'commentary: evaluation' for **this** in lines [52] to [54].

[52] ... This holds true according to The Master Switch ...

[53] ... this is the quintessential nature of the African system ...

[54] ... this is far from ideal ...

5.2.1.6 Setting things in interpretive contexts

Durrant finds one of the final distinguishing features of student Humanities and Social Sciences writing within his study to be the focus on ‘setting [things] in interpretative or explanatory context’ (2015, p. 24). In the current study this interpretative function is found within phraseologies for **that**, **as** and **this**.

Within phraseological categories for **that**, interpretation is central to phraseologies categorised as possessing the semantic motif ‘voice of the essay writer – interpretation’ (illustrated in lines [55] to [57]).

[55] ... The complexity of the relationship to Soviet Russia **demonstrates that** ...

[56] ... **This illustrates that** Bayart’s thesis is very relevant in understanding ...

[57] ... **There is more evidence to suggest that** there was a difference between servants and the ...

The ‘interpretation’ motif which is the largest semantic category for **as** also performs this function (illustrated in lines [58] to [60]).

[58] ... **interpreted as** imperialist by Mao ...

[59] ... **we should look at** the United Nations system objectively **as** a forum for nations to come together and tackle issues ...

[60] ... **we cannot view** Pompadour **as** having been popular in Versailles’ ...

Interpretation is also the central function of phraseologies within the ‘commentary: interpretation’ motif for **this** (illustrated in lines [61] to [63]).

[61] ... **This is perhaps why** the creation of the hydrogen bomb seemed ...

[62] ... **this does not necessarily indicate** the complexities ...

[63] ... **this may reflect concern about** prosecution of dissenting states’ citizenry ...

5.2.1.7 Conclusions

Sections 5.2.1.1 to 5.2.1.6 have examined the phraseologies within the current study which exhibit continuity across the four disciplinary/institutional sub-corpora for how these continuities may or may not be connected to 'discipline' in terms of characteristics shared at the level of what could be called 'meta-disciplinary' groupings. For this purpose, an attempt has been made to map features of the phraseologies with features identified by Durrant as being characteristics specific to student Humanities and Social Sciences writing in contrast to student Science and Technology writing.

This mapping process has been by no means exhaustive; it has largely concentrated only on the most salient and obvious examples of how the phraseologies link to each Humanities and Social Sciences feature. A more comprehensive examination of phraseologies across keywords and semantic categories would be likely to yield even more such links.

A picture has been built up of wide-ranging links between the functions of the phraseologies across all grammatical keywords and the broad purposes identified by Durrant (2015) as distinctive to student Humanities and Social Sciences writing. This suggests that discipline at the meta-disciplinary level has a strong relationship to the similarities identified across the four disciplinary/institutional sub-corpora. It can be argued that the current study offers a new angle on how the purposes identified by Durrant can be realised linguistically beyond lexical bundles.

5.2.2 Variations

Findings of variation make up a much smaller proportion of the results than do findings of continuity. Most of this variation, and, indeed, all variation that can be linked to discipline, is located within the findings for *this*. In this section the findings

for *of*, *and*, *that* and *as* will first be discussed together, and then findings of variation for *this* will be examined separately.

5.2.2.1 *and*, *of*, *that* and *as*

None of the apparent variations in the findings regarding frequency of particular phraseologies for *and*, *as*, *of*, and *that* can be categorised as either disciplinary or institutional. Potential differences in frequencies of particular phraseologies between two disciplines at one institution were not replicated across both institutions. Potential disciplinary or institutional differences identified in the analysis of the first 300-line random samples as occurring across all four sub-corpora were also not replicated in the subsequent analysis of second 300-line random samples.

These two observations serve as important validation for two aspects of the research design in the current study. Firstly, had the study not included an institutional dimension, false conclusions may have been drawn with respect to the influence of discipline; for example, if only HIST-Y and POL/IR-Y sub-corpora had been compared, the high incidence of ‘reason’ phraseologies for *as* discussed above would have wrongly been attributed to discipline. Secondly, the fact that analysis of second random samples often did not support differences identified from the first samples shows the importance, in the context of this study at least, of the use of second samples to verify any potential findings.

While there are no clear disciplinary or institutional variations for *of*, *and*, *that* and *as*, there are variations concerning individual sub-corpora. The most notable examples of individual sub-corpus variation are the relatively high use of the fixed ‘exemplification’ phrase *such as* within the HIST-X sub-corpora, the relatively high incidence of phraseologies with the semantic motif ‘reason’ within the POL/IR-Y sub-corpora findings for *as*, and for *that* the higher proportion of ‘voice of the essay writer’ in contrast ‘voice of others’ phraseologies in the POL/IR-Y sub-corpora which

is the reverse of the relative ‘voice’ ratios in the other three sub-corpora (N.B. in the findings for **of** all individual variations were smaller and consequently the reasons for these differences were not examined). As previously discussed within the Results section, these differences can be linked to over-representation of a particular student author or particular student authors within the phraseologies for a particular semantic category and are thus best explained by individual writer idiosyncrasies.

5.2.2.2 **this** - overview

In contrast to the lack of clear disciplinary or institutional differences in the findings from analysis of **as**, **of** and **that**, in the analysis of **this** seven differences along disciplinary lines were found which were all verified by analysis of second random samples. Table 160 below summarises these differences.

| | Higher frequency in History | Higher frequency in Politics/International Relations |
|---------------|---|--|
| Encapsulation | <ul style="list-style-type: none"> ▪ ‘Encapsulation as <i>object</i>’ | <ul style="list-style-type: none"> ▪ ‘Encapsulation as <i>the essay text</i>’ ▪ ‘Encapsulation as <i>scholarly angle</i>’ |
| Commentary | <ul style="list-style-type: none"> ▪ ‘Commentary: <i>addition</i> to preceding statement’ ▪ ‘Commentary: <i>cause</i> of preceding statement’ | <ul style="list-style-type: none"> ▪ ‘Commentary: <i>meaning</i> of preceding statement’ ▪ ‘Commentary: <i>exemplification</i> of preceding statement’ |

Table 160: Disciplinary differences in the phraseologies of **this**

It should be noted that in all cases they are a matter of degree: there is a higher instance of a particular semantic category in one set of disciplinary sub-corpora, suggesting a greater preference for the function it performs; however, this function is not completely absent from the sub-corpora representing the other discipline:

- The semantic category ‘encapsulation as object’ made up approximately 13% to 16% of the History sub-corpora random samples across both first and second samples, compared to approximately 10% to 11% in the Politics/International Relations samples.

- The semantic category ‘commentary – interpretation’ made up approximately 11% to 14% of the Politics/International Relations sub-corpora random samples across both first and second samples, compared to approximately 6% to 8% in the History samples.
- The semantic category ‘encapsulation as the essay text’ made up approximately 8% to 11% of the Politics/International Relations sub-corpora random samples across both first and second samples, compared to approximately 2% to 4% in the History samples.
- The semantic category ‘encapsulation as scholarly angle’ made up approximately 9% to 13% of the Politics/International Relations sub-corpora random samples across both first and second samples, compared to approximately 6% to 9% of the History samples.
- The semantic category ‘commentary – addition’ made up approximately 5% to 7% of the History sub-corpora random samples across both first and second samples, compared to approximately 4% to 5% of the Politics/International Relations samples.
- The semantic category ‘commentary – exemplification’ made up approximately 3% to 5% of the Politics/International Relations sub-corpora random samples across both first and second samples, compared to approximately 1% of the History samples.
- The semantic category ‘commentary – cause’ made up approximately 2% of the History sub-corpora random samples across both first and second samples, compared to approximately 0% to 1% of the Politics/International Relations samples.

We are thus talking about *greater levels of emphasis* on particular semantic functions for **this** in the different disciplinary sub-corpora. The following sections will attempt to explain these differences in emphasis in student writing in the two disciplines first examining the ‘encapsulation’ motifs then the ‘commentary’ motifs.

5.2.2.3 Greater emphasis on ‘object’ vs. greater emphasis on ‘the essay texts itself’ and ‘scholarly angle’

There is more emphasis on encapsulating objects in History and there is more emphasis on encapsulating the essay text itself and scholarly angles in Politics/International Relations. A useful tool for explaining this difference in emphasis is the framework Gosden (1993) developed in his analysis of the discourse function of the grammatical subject in scientific RAs. He investigated the balance ‘between interactional ‘human face’ discourse and impersonal topic-based discourse’ (p. 56) and outlined four main *subject role domains*, the Participant domain, the Discourse domain, the Hypothesized and Objectivized domain and the Real World domain (p. 62). Gosden’s model of these domains is reproduced in Figure 6 below.

| | PARTICIPANT DOMAIN | DISCOURSE DOMAIN | HYPOTHESIZED AND OBJECTIVISED DOMAIN | REAL-WORLD DOMAIN |
|--|---|--|--|--|
| <p>↑</p> <p>MORE EXTERNAL COMMUNITY ORIENTATED THEME</p> | <p>INTERACTIVE PARTICIPANT (Smith 1987)</p> | <p>INTERACTIVE DISCOURSE ENTITY (previous studies)</p> <p>EMPTY DISCOURSE THEME (it is concluded) (there are reports)</p> <p>Micro DISCOURSE ENTITY (Figure 1b)</p> <p>Macro DISCOURSE ENTITY (this paper)</p> <p>DISCOURSE EVENT/PROCESS (the conclusion)</p> | <p>HYPOTHESISED ENTITY (the model)</p> <p>EMPTY H AND O THEME (it was evident) (There was evidence)</p> <p>OBJECTIVISED VIEWPOINT (one factor)</p> <p>HYPOTHESISED VIEWPOINT (the possibility)</p> | <p>REAL-WORLD ENTITY (AlFeNi alloy)</p> <p>EMPTY REAL-WORLD THEME (It was found) (There was a delay)</p> <p>REAL-WORLD EVENT/PROCESS (Annealing)</p> <p>MENTAL PROCESS (Deduction)</p> |
| <p>↓</p> <p>MORE INTERNAL WRITER-ORIENTED THEME</p> | <p>PARTICIPANT VIEWPOINT (Our argument)</p> <p>DISCOURSE PARTICIPANT (We)</p> | | | |
| | <p>MORE INTERACTIONAL THEME ←-----→ MORE TOPIC-BASED THEME</p> <p>↕</p> <p>GREATEST WRITER VISIBILITY</p> <p>↕</p> <p>LEAST WRITER VISIBILITY</p> | | | |

Figure 6: Model of Subject role domains (Gosden, 1993, p. 63)

Gosden's model shows the domains along a cline from 'greatest writer visibility' to 'least writer visibility' with foci within these domains organised on a cline from 'more external community oriented' to 'more internal writer-oriented' themes. Using this model, it can be argued the 'encapsulation as object' motif within the current study is closely aligned to Gosden's 'Real-world domain'. In contrast, the 'encapsulation as the essay text itself' motif maps to the 'Macro discourse entity' element within Gosden's 'Discourse domain'. This is also the domain in which 'encapsulation as scholarly angle' would fit; however, because phrases with the 'encapsulation as scholarly angle' semantic function tend to express voices other than the essay writer, they would generally sit at the 'more external community oriented' as opposed to the 'more internal writer-oriented' end of the 'theme' cline.

Thus, using Gosden's model, it can be argued that the differing emphases in frequency of different encapsulation motifs for *this* within History and Politics/International Relations suggest a greater emphasis on the real-world domain of the particular field of focus in History, in contrast to a greater emphasis on the academic discourse occurring in relation to the field in focus within Politics/International Relations. Within the Politics/International Relations sub-corpora, the essay itself is made more visible as a 'participant' in this discourse suggesting a possibly greater overall level of writer visibility in Politics/International Relations.

The differences in 'domain' focus outlined above reflect certain of the differences outlined in section 3.5.2 of Chapter 3 in terms of the focus of the essays texts in the two disciplinary groupings across both institutions as inferred from an analysis of the essay questions. The majority of essay questions within the History sub-corpora entailed a focus on *the specific* - specific actors, specific locations, specific contexts and specific events - as opposed a focus *general* concepts or ideas, issues and theories. It therefore makes sense that there would be greater frequency of encapsulation of 'objects' within the History texts. Indeed a quick count across the first random 300-line samples for the four disciplinary/institutional sub-corpora

shows a higher number as well as a higher proportion of ‘object’ elements referring explicitly to aspects of the real-world field in focus in History: in the HIST-Y sub-corpora within the ‘encapsulation as object’ motif we get *immense historian, extraordinary woman, painting (x2), newcomer, exhausted country, document, bourgeoisie, passage, prison*; in the HIST-X sub-corpora within the ‘encapsulation as object’ motif we get *criminal class, guild, Christian bishop, British outpost, letter, type of medicine, play (x2), and fashionable commercial spot*. In comparison, this type of focus is numerically and proportionately less frequent within the smaller number of ‘encapsulation as object’ phraseologies within both Politics/International Relations sub-corpora. With only two occurring in POL/IR-Y, *low income group* and *family*, and only four occurring in POL/IR-X, *small elite, underclass, nation* and *region*.

Similarly, the greater focus on ideas and theories in the essay questions in the Politics/International Relations sub-corpora means it makes sense that there would be a higher relative frequency of ‘encapsulation as scholarly angle’ phraseologies within the Politics/International Relations sub-corpora. This focus on ideas and theories also means an automatically greater focus on the discourse happening ‘around’ real-world phenomena which can be argued to help explain the greater use of ‘encapsulation as the essay text itself’ phraseologies as well: for the essay-writer in Politics/International Relations there is a greater focus on engaging with ‘a discourse’ rather than representing the processes of real-world phenomena (the greater focus required in History essays), and, it therefore makes sense that the essay itself would be more likely to be portrayed as a ‘participant’ in this discourse.

5.2.2.4 Greater emphasis on ‘addition’ and ‘cause’ vs. greater emphasis on ‘interpretation’ and ‘exemplification’

There is more emphasis on adding to ideas and explaining causes of phenomena in History and there is more emphasis on interpreting and exemplifying phenomena in Politics/International Relations. To an extent, these differences can also be explained by differences with regards to the focus of the essay texts as inferred

from surveying the essay questions in the two disciplinary groupings (again, see section 3.5.2 of Chapter 3). In the disciplinary/institutional sub-corpora causality is a much greater focus in History than in Politics/International Relations so it therefore makes sense that cause-focused phraseologies would be more frequent in the History sub-corpora. Focus on specificity is also more prevalent in History which could be argued to link to a slightly greater frequency in the sub-corpora for this discipline of addition-focused phraseologies which help to build up a detailed picture. This is illustrated in lines [64] and [65] below both of which perform the function of simply adding further detail to an idea.

[64] ... **this** took place in the higher echelons of medical practice ... [HIST-X]

[65] ... **this** was redistributed when necessary ... [HIST-Y]

In contrast, in Politics/International Relations more focus on general concepts applicable across a range of specific contexts can be argued to help explain the higher frequency of exemplification phraseologies which provide specific examples for general ideas or arguments, as seen in lines [66] and [67].

[66] ... **This** is evident in the former Yugoslavia ... [POL/IR-Y]

[67] ... **This** has been the case in the 25 year Sri Lankan civil war ... [POL/IR-X]

The greater emphasis on 'interpretation' commentary phraseologies is a bit more difficult to neatly explain in relation to disciplinary differences in essay focus. However, this could be very tentatively linked to the suggestion made above that, in Politics/International Relations, there is a relatively greater explicit emphasis on the discourse occurring in relation to the field in focus as opposed to focus on the real-world entities and processes within the field of focus. This emphasis could be tentatively argued to mean that, in contrast to student History writing, there is

more ‘talk’ in student Politics/International Relations writing about ‘what things mean’, as illustrated in lines [68] and [69] below.

[68] ... **this** reflects unrealistic expectations of the UN as an actor ...
[POL/IR-X]

[69] ... **this** does not necessarily translate to an inevitable conflict ...
[POL/IR-Y]

5.2.2.5 Conclusions

Sections 5.2.2.1 to 5.2.2.4 have examined the smaller number of phraseologies within the current study which exhibit variation. Section 5.2.2.1 outlined variations in individual sub-corpora, the most noticeable to be found within particular phraseological categories for *that* and *as*, and argued that these can largely be explained in terms of individual writer idiosyncrasies.

Sections 5.2.2.2 to 5.2.2.4 outlined variations which fall along disciplinary lines in the degree of relative emphasis on particular ‘encapsulation’ and ‘commentary’ phraseologies for *this*. It is important to reiterate here the *relative* rather than *absolute* nature of these differences. All ‘encapsulation’ and ‘commentary’ motif categories are present in both disciplines, and, indeed, many have already been discussed in the explanation of continuities across History and Politics/International Relations in section 5.2.1 above.

Gosden’s (1993) framework of subject role domains falling along a cline from more interactional, involving greater writer visibility, to more topic-based, involving least writer visibility, has been argued to be useful in explaining the possible reasons for disciplinary differences in emphases regarding frequency for particular *this* ‘encapsulation’ phraseologies. The greater frequency of ‘encapsulation as object’ phraseologies in History suggests greater relative focus on the real-world domain

within student writing in History. In contrast, the greater frequency of 'encapsulation as the essay text' and 'encapsulation as scholarly angle' phraseologies within Politics/International Relations suggest greater relative focus on the discourse domain linked to a possibly higher level of writer visibility within student Politics/International Relations writing.

The different patterns of focus according to discipline in the essay questions, as outlined in section 3.5.2 of Chapter 3, has also been argued to have use in explaining the possible reasons for disciplinary differences in frequency of both particular 'encapsulation' and as well as particular 'commentary' phraseologies for *this*. A greater focus on specificity in History links to relatively greater use of 'encapsulation as object' and 'commentary: addition' phraseologies, and a greater focus on causality links to greater use of 'commentary: cause' phraseologies. A greater focus on ideas and theories within Politics/International Relations links to a greater focus on 'encapsulation as scholarly angle' phraseologies, and, more focus on general concepts in Politics/International Relations links to greater frequency of 'exemplification' phraseologies providing examples of specific phenomena which serve to illustrate general ideas.

More tentatively, greater frequency of 'encapsulation as the essay text itself' phraseologies has been linked to relatively greater explicit focus on the discourse domain around real-world themes in Politics/International Relations and its likely link to a greater sense of the essay writer as 'participant' in this discourse. Even more tentatively, it could be argued that greater frequency of 'commentary: meaning/interpretation' phraseologies within Politics/International Relations may be linked to relatively greater focus on the discourse occurring in relations in that discussion of 'what things mean' links more strongly to a discourse domain focus than a real-world field focus.

In light of the examination above of similarities and variations, it is also worth commenting on a contrast between the results of the current study of History and

PIR writing and the results of Groom's (2007) comparison of professional writing in the disciplines of History and Literary Criticism. In the current study, the number of similarities in semantic sequences and motifs used in the writing in both target disciplines very strongly outweighed the number of variations between them. By contrast, in Groom's study, whilst again there were a large number of continuities across the two target disciplines, there were also a large number of differences. This suggests that the writing in History and PIR sits closer on a disciplinary cline than History and Literary Criticism. This is interesting considering the fact that within traditional disciplinary meta-grouping practices, History and Literature tend to at least look as if they are the closer-sitting disciplines. History and Literature sit more firmly within Becher and Trowler's (2001) 'soft pure' disciplinary category; PIR would likely be seen as a more tenuous member of this category, very close to being 'soft applied'. Moreover, within Nesi and Gardner's (2012) disciplinary model whilst History and Literature both lay within the 'Humanities' grouping, PIR fell within the 'Social Sciences' grouping. However, the stronger relationship at lexico-grammatical level between History and PIR compared to History and Literature is, as discussed above in 5.2.1, reflected in Durrant's (2015) grounded study of lexical bundles which puts History, Politics and Law in a tight sub-cluster within a wider Humanities and Social Sciences grouping. Thus, it could be argued that the differences between the current study's results and Groom's results (2007) provide further evidence that, when considering disciplinary discourse, no assumptions should be made about shared features based on conventional disciplinary meta-groupings or pre-existing models that have not been drawn from linguistic evidence; rather, further grounded studies of disciplinary language and discourse such as that conducted by Durrant (2015) are needed.

5.3 The influence of discipline vs. institution

Section 5.2 has outlined and attempted to explain both the continuities and the variations in semantic motifs and sequences within phraseologies for *of*, *and*, *that*,

as and **this**. Although continuities are the dominant pattern in the findings, there are also clear patterns of difference across the sub-corpora most noticeably for the keyword **this**, and both continuities and variations can largely be explained in terms of discipline. As has been discussed in section 5.2.1, continuities in the semantic patterning of grammatical keywords can to a great extent be explained in terms of the shared features of the meta-disciplinary grouping of student Humanities and Social Sciences writing (Durrant, 2015). Similarly, as discussed in section 5.2.2, the findings of difference, apart from instances of individual writer idiosyncrasy, can also be largely explained in terms of preferences of the individual disciplines of History and PIR with there being no cases of clear differences falling along institutional rather than disciplinary lines.

Thus, the findings for the two disciplines and two institutions that constitute the focus of the current study suggest that students' writing is more likely to be influenced by regularities in the practices of a particular discipline or regularities in the shared practices of a meta-disciplinary grouping across institutions and departments than it is to be strongly influenced by the idiosyncrasies of an individual institutional context. In the current study at least, discipline seems to override institutional context in its influence on the nature of student academic writing in History and PIR at the lexico-grammatical level.

5.4 Contributions to knowledge & implications

Having discussed the results themselves, it is now time to assess the contribution of this project as a whole to its field.

The findings of this project contribute to the picture that is being built up in the still relatively nascent body of literature concerning student academic writing across the disciplines. As discussed in Chapter 2, until very recently, research with a focus on the disciplinary features of student texts was rare. At the macro cross-disciplinary

level, Nesi and Gardner's (2012) study has been seminal in terms of our understanding of the purposes of undergraduate and taught postgraduate coursework assignments and the genres realising these purposes across the academy. Also at the macro cross-disciplinary level, studies by Hardy and Römer (2014) and Durrant (2014, 2015) have provided us with surveys of the continuities and variations to be found in the linguistic features of student texts across disciplines, with the former research involving Biber-style dimensional analysis and the latter involving focus on vocabulary and lexical bundles respectively. At the more 'micro' level of close research into student writing in particular disciplines or pairs or small groups of disciplines, some studies have involved contrastive analyses with a disciplinary focus (e.g. Gardner 2008, 2012, Thompson, 2009, Bruce, 2010, Li and Wharton, 2012), but this focus has been relatively rare; a larger majority of studies using the BAWE, for example, involve contrasting learner with proficient writing or student with professional writing.

The current study has involved a close bottom-up analysis and comparison/contrast of phraseologies within the same student genre family, Essays, across the two 'soft' disciplines of History and Politics/International Relations. To my knowledge, at the time of writing, there have been no other studies conducted with a specific close focus on UK student Politics/International Relations writing. A focus on History has been a feature of two other recent investigations of UK student writing: Gardner (2008) conducted an Angle on Field (Martin, 1993) analysis of sentence subjects within BAWE texts and included History as part of this analysis, and Thompson (2009) looked at 'it + *linking verb* + ADJ + that' and 'it + linking verb + ADJ + that' sequences, frequent n-grams and the use of personal pronouns within BAWE History texts in contrast to BAWE Engineering texts. The current study analyses student History texts from a new perspective, that of semantic patterning of grammatical keywords, and, in restricting the analysis to Essay texts, also provides, to my knowledge the first detailed genre-specific contrast of the features of UK student History writing with student writing in the same genre family from another discipline.

However, in such a sparsely populated field, it is easy enough to make a ‘new’ contribution with regards to focus. The real value of a contribution needs to be measured in terms of how much it reveals about the nature of student writing in the disciplines, and in terms of its potential usefulness pedagogically and as a research tool in future studies.

5.4.1. Coverage: operationalizing the text

As already discussed at the beginning of Chapter 4 (see section 4.1), a strong argument can be made for the usefulness of the findings of the current study in terms of ‘coverage’ in the sense of the proportion of the target texts the analytical tool used reaches. It is worth reiterating here Bruce’s (2011, p. 6) argument that pedagogical usefulness ‘rests on the extent to which ... researched features of academic texts, of themselves, can be said to operationalize the wider phenomenon of academic subject discourses realised in texts’; these features need to occur frequently enough within target texts to be useful to teach.

The five target grammatical keywords, *of*, *that*, *as*, *this* and *and*, on their own make up approximately 10% of the texts across the four disciplinary/institutional sub-corpora, and, as described at the beginning of Chapter 4 (see section 4.1), if all phraseologies which had been semantically categorised as part of my analysis as occurring around these five keywords in one randomly-selected essay text (PIR3_008_R from POL/IR-X) are highlighted, this results in coverage of more than 70% of the text (see Appendix 3). This is in stark contrast to my second analysis of the same essay (also to be found in Appendix 3) taking three examples of other tools used for a lexico-grammatical discourse focus in studies of BAWE texts: Thompson’s (2009) it + *linking verb* + ADJ + that and it + *linking verb* + ADJ + to phrases and focus on the first person pronoun (also a focus in Nesi and Gardner (2012)); Nesi and Gardner’s (2012, p. 118) IF-THEN sequences; and lexical bundles in Durrant’s study (2015, pp. 17-18) which have been categorised as ‘[d]istinctive Humanities and Social Sciences bundles’. I manually searched the same essay text for all the above features and found either very few or none of each. There were no

examples of it + *linking verb* + ADJ + that or it + *linking verb* + ADJ + to phrases (I did find one 'It ... that' phrase, *it can be seen that*, which has been put in bold red italic font in the essay text in Appendix 3), no examples of IF-THEN sequences, only two instances of the first person singular 'I' (in bold red italic font in the text in Appendix 3), and two instances of Humanities and Social Sciences lexical bundles, *the power of the* and *the way in which* (in red bold italic font in the text in Appendix 3).

In the respective studies of student writing in which they were used, the above analytical tools have proved useful for revealing disciplinary preferences and tendencies across multiple texts in the target discourses. However, for the purposes of understanding the workings of a single text from a particular discourse and genre, these tools prove to be of extremely limited use. The findings in these studies can therefore be argued to have limited 'operationalization' value for pedagogical purposes in the sense described by Bruce (2011, p. 6).

In contrast, the analytical tool used within the current project, the semantic patterning of grammatical keywords, has arguably proved useful for understanding disciplinary features of student writing both across multiple texts and within individual texts. The findings have contributed to a 'meta' cross-textual understanding of repeated features and tendencies within the target discourses as a whole, as well as to a rich, detailed understanding of what goes on within individual texts within these discourses. There is therefore a high 'operationalization' (Bruce, 2011) value and consequently fruitful pedagogical implications which will be discussed in more detail below.

5.4.2 The role of language: Language as academic purpose

The findings of the current study shed light on the role of language within the target student disciplinary discourses. A wide range of disciplinary functions, such as a focus on the abstract, depiction of multiple contingent viewpoints, and evaluation, are realised in student texts by the frequently used phraseologies which, as we have

seen, constitute a very considerable proportion of these texts. It can be argued that these phraseologies are, in many ways, 'doing' the disciplinary activity of student writing in History, 'doing' the disciplinary activity of student writing in PIR, and 'doing' many of the shared disciplinary activities of student Humanities and Social Sciences writing.

This observation is in line with Turner's (2004) argument that language is constitutive of academic purpose. Turner critiques the notion of a language/content dichotomy in which language plays a subordinate role as an 'instrument' in which proficiency needs to be fine-tuned then applied, as a finished product, to the higher content purposes of academic activity. She argues (p. 99):

The failure to see that language grows with content leads to the insidious perception that measured language proficiency is all there is to the language on a degree course ... [this is] a gross misunderstanding of the nature of language, its constitutive importance in academic performance, and the intellectual challenge of learning its uses and using it well.

Turner goes on to argue that there is a tendency in EAP to privilege focus on 'macro- or structural levels of organisation' over attention to lexical and phraseological accuracy and that 'because fluency has won in EAP, we are over-lenient on our students' mistakes' (p. 100). She argues (p. 105):

Understanding the language thoroughly, as well as why it has been used in the way it has been used, heightens understanding of the ways of thinking in a discipline.

In her article Turner does not explicate how exactly language realises academic or disciplinary purpose, but it could be ventured that the current study offers one illustration of this linguistic realisation, specifically for the disciplines of Politics and History at undergraduate level, and thus provides concrete support for Turner's 'language as academic purpose' argument.

If we accept the 'language as academic purpose' argument, then we accept that learning and effectively using specific-purpose language is essential to the successful 'doing' of an academic discipline both at professional and at student

level. It is therefore crucial to understand what acquisition of specific-purpose language involves. Gee's (1989) concept of "secondary discourses" is helpful in this respect.

Gee argues that beyond our "primary discourse", the 'socio-culturally determined ways of using our native language in face-to-face communication with intimates', are "secondary discourses", 'which crucially involve social institutions beyond the family' (p. 22). He suggests that a useful way to define literacy is as 'control of secondary uses of language (i.e., uses of language in secondary discourses)'. This idea of secondary discourse control as literacy relates to the way *academic* literacy is conceptualised by Rose, Rose, Farrington and Page (2008) who see it as the accessing of technical and/or social science discourses which enable the reading of high level academic texts and successful use of this reading in academic writing.

With regards to developing literacy, Gee makes the distinction between acquisition, 'the process of acquiring something subconsciously by exposure to models and a process of trial and error', and learning, 'conscious knowledge gained through teaching', of secondary discourses (1989, p. 20). He argues that acquisition is ideal and that mainstream middle-class students are advantaged in school because it gives them the opportunity to practice the secondary discourses they have already begun the process of acquiring in their homes. In contrast, non-mainstream children, who lack access to secondary discourses at home, 'cannot practice what they haven't yet got' and are thus 'exposed mostly to a process of learning and not acquisition'.

Rose et al. (2008) describe in similar terms the problem that many indigenous adult university students in Australia face with regard to developing academic literacy. These students have not experienced the secondary school apprenticeship in reading, writing and discussing the academic discourses and, thus, '[a]t average rates of development it could take potentially up to six years for these students to acquire the academic literacy necessary for tertiary study' (p. 167). Rose et al. (2008) argue that the academic literacy challenge faced by indigenous Australian adult

university students is also experienced, for different reasons, by many international students (although having received a full secondary school education and therefore likely developed the equivalent academic literacy within their own language and culture, international students studying within an English-language academic context need to then rebuild this academic literacy in English). Rose et al. (2008) offer as a remedy to this acquisition gap the *Scaffolding Academic Literacy* pedagogy which is designed to accelerate the learning and acquisition of academic discourses (p. 165) through the integration of the dimensions of the academic field and the 'language patterns in which it is expressed' (p. 167) through the supporting both the academic reading process and the academic writing process. Attention to language patterns in the reading texts and the modelling of language patterns for writing is central to this pedagogy.

From a 'language as academic purpose' perspective, with regards to pedagogical implications, the principles of the Scaffolding Academic Literacy approach chime with the findings of the current study. The semantic patterning of grammatical keywords has been shown to play a central role in the constitution of linguistic patterns through which the two academic fields of History and Politics/International Relations are expressed. Gee's distinction between acquisition and learning are pertinent here: it is possible the competent third-year student authors of the texts in this study acquired rather than learned the phraseologies of their discipline, developing and fine-tuning disciplinary competence over the course of their degree programmes. However, the phraseologies identified within the current study represent a useful starting point for accelerated acquisition through the teaching of language patterns constitutive of discipline.

5.4.3 The nature of disciplinary language & pedagogic implications

The findings of the current study contribute to understanding the nature of disciplinary language, at the level of meta-disciplinary groupings and at the level of individual disciplines, in student texts. A phraseological perspective which sees

academic language in terms of the semantic patterning of grammatical keywords is an extremely useful way of understanding the student History and Politics/International Relations texts at the centre of this research. This approach has facilitated the building of a fairly comprehensive taxonomy of functions and meanings being performed in these texts, what student writers of the texts are 'doing' - arguably the disciplinary activity of the texts - and how this is realised linguistically. The extensive coverage of the texts that this way of seeing them affords provides further strong support for the usefulness of this phraseological perspective.

If we consequently accept that the argued usefulness of this research approach can potentially be translated into usefulness pedagogically, then this would lead to the contention that, in the case of the two target disciplines at least, an approach with a heavy lexical emphasis, as opposed to an approach foregrounding grammar, would prove the most effective way of teaching this type of academic language and discourse. The phraseologies entail strings of language which are best understood first functionally, then lexically, then grammatically. Examples of the semantic patterning of **that** in lines [70] to [78] below, taken from the first pooled sample, can be used to illustrate this argument.

[70] ... They asserted **that** reforming the system and emphasising good governance ...

[71] ... Marxism and Anarchism both hold **that** power works downwards ...

[72] ... Mark Freeman argues **that** historians tend to rate such journalism as ...

[73] ... It has also been argued **that** this entails that all subsequent writings ...

[74] ... Another argument against ad-hoc tribunals is **that** they are expensive and ...

[75] ... an interesting suggestion **that** states are becoming more uniform as a result of ...

[76] ... it is clear **that** contemporaries held little hope of ...

[77] ... It is important to note **that** in the moral debates about luxury ...

[78] ... The complexity of the relationship with Russia demonstrated **that** the character of the revolution cannot be ...

To summarise what was argued in Chapter 4, the set of **that** phraseologies above have the semantic function of expressing ‘voice’, either the voice of the essay writer or the voice of ‘others’, and the elements preceding **that** work to create an angle that frames, in a particular way, the statement or proposition following **that**. This is the best place to start in teaching the use of **that** within academic essays, for the disciplines of History and Politics/International Relations at least. A lesson focusing on **that** might start by raising students’ awareness the fact that the bit of language preceding **that** is used by the essay writer for the purpose of communicating to the reader *how* they should understand the statement following **that**.

Following on from this initial focus on semantic function, there should then be focus on how the various phraseological elements that make up the bit preceding **that** contribute to this purpose. The bit of the phraseology preceding **that** can include elements expressing **AUTHOR OF STATEMENT**, *they, Marxism and Anarchism, Mark Freeman*, **CHARACTER OF STATEMENT**, *asserted, both hold, argues, has also been argued, Another argument against ad-hoc tribunals, an interesting suggestion*, **ESSAY WRITER ANGLE ON STATEMENT**, *is clear, is important to note*, **PHENOMENON AS EVIDENCE**, *The complexity of the relationship with Russia*, and **INTERPRETATIVE ANGLE**, *demonstrated*. An exploration of these elements would involve attention to the sequence of words realising them, and this would include focus on the lexical sets strongly associated with these element-forming sequences, for example, the various verb lemmas used within the **CHARACTER OF STATEMENT** element such as *state, argue* and *claim*.

It is at this stage that attention to grammatical form would become useful: whether the lemma *argue* is used in present-tense active (as in line [72]) or present perfect passive verbal form (as in line [73]), or in a nominalised form (as in line [74]); if the reporting verb is nominalised, whether that means the element it is constitutive of becomes a clause subject, followed by the verb to be, and **that** plus the element following it becomes the sentence complement, or, whether it means that the whole phraseology for **that** becomes a nominalised subject; and, what these two choices mean within a wider stretch of the text. Thus, in a semantically-driven phraseological approach to teaching academic discourse, a focus on grammar could be argued to occupy a subservient level as a ‘mechanical’ service to the lexical sequences realising semantic functions within a text. Such an approach, going from purpose, to phraseology and lexis, to grammar, is in line with the argument made by Durrant and Matthews-Aydinli (2011) for the usefulness a ‘function-first’ as opposed to ‘form-first’ approach in identifying formulaic language within academic texts.

Whilst it could be argued that, in so far as the phraseologies are patterned around grammatical keywords, this approach does have ‘grammar’ at its heart, the way in which these grammatical words are being approached here is not primarily to do with the range of their structural functions but with the semantically-driven behaviours of words that occur with them in specialised discourses. The potential pedagogical approach described above links to Willis’s argument for a Lexical Syllabus which rests on the idea that ‘word meaning and word order are central to English in a way that may not hold true for the other languages’ (1990, p. 24).

In terms of the nature of disciplinary language, a prominent structural feature also deserves attention here. The analysis in this project has been semantic not grammatical, and, furthermore, I departed somewhat from Groom’s (2007) approach by foregoing the structural analysis and categorisation phase for each grammatical keyword, which he conducted before his semantic categorisation of phraseologies. However, it has been impossible to ignore, and, indeed, it would be foolish to ignore, the high frequency of one particular structural property across

phraseologies and keywords in this study, that of nominalisation. Across all four sub-corpora within the study approximately 90% of the phraseologies for *of* are nominalised. More than 20% of *that* phraseologies consisted of nominalised **PHENOMENON + that + DETAIL** phraseologies (see line [80]), and a sub-grouping of the ‘the voice of others’ motif was the semantic sequence **CHARACTER OF STATEMENT + that + STATEMENT** (see line [81]). Between 45-52% of phraseologies for *this* were ‘encapsulation’ phraseologies which are nominalised. The data for *as* included fewer nominalised phraseologies, but there were a few within the ‘function’, ‘role’ and ‘conceptualisation’ motif groupings (line [83] comes from the ‘conceptualisation’ grouping in the pooled sample).

[79] ... the overriding geopolitical tensions **of** the early Cold War ...

[80] ... the political and theoretical context **that** led to the 2003 invasion of Iraq ...

[81] ... the idea **that** society can and does provide equality ...

[82] ... **This** shift in population and employment patterns ...

[83] ... economics **as** a factor ...

This finding is in line with Groom’s (2007) finding of a prevalence of nominal structures in professional academic writing in History and Literary Criticism which he observed was ‘entirely consistent the canonical view of academic written English across the disciplinary spectrum as highly nominalized in nature’ (p. 252). It also supports Schleppergrell’s argument (2004a, 2004b) that nominalisation features prominently in successful student writing.

Martin provides a useful explanation of the essential role of nominalisation, for the purposes of ‘technicality’ in the Sciences and for the purposes of ‘abstraction’ within the Humanities, in the construction of disciplinary fields in written academic discourse (1993, p. 228):

Both technicality and abstraction depend on the same linguistic resource, nominalization ... Sound is not a thing, but has to be dressed up as one in

scientific discourse in order to be defined. Similarly in history, realizing reasoning inside rather than between clauses means placing an Agent in causal relation to its Medium, and this entails nominalizing events as participants and verbalizing the logical relationship between them. The enlargement of Australia's steel-making capacity and the demands of war are not things, any more than sound is, but they have to be grammaticalized as things in order to reason within the clause. And this underlines the significance of writing in the development of discipline – specific discourses – grammatical metaphor is primarily a resource for writing, not speaking. ... Without the technology of writing, science and history as we practice them would not exist.

Parkinson and Musgrave (2014) investigated development of noun phrase complexity in student writers' texts by comparing use by students preparing for academic study and students already enrolled in graduate study, and identified stages in the development of noun phrase use that were largely in line with previous findings (e.g. Biber and Gray, 2011). They suggest that 'to speed up students' acquisition of a more academic register' focus the more complex noun phrase constructions would be useful (p. 58).

This point brings us back to the issue, discussed earlier in this section, of what is the most appropriate pedagogic approach to development of discipline-specific academic language and discourse. Parkinson and Musgrave (2014) argue for structurally-focused attention to the more complex nominalised constructions, such as those involving abstract meanings and post-modification with prepositional phrases, '**the production of fossil fuels**', or those involving complement clauses controlled by nouns, '**viewpoint that using nuclear energy is equal to suicide**' (p. 50). These two examples from Parkinson and Musgrave would have been discussed within the current study as the semantic sequences **PROCESS + of + OBJECT** and **PHENOMENON + that + DETAIL** respectively. I would argue that it is most useful pedagogically to see development of noun phrase complexity as evidence of development of the ability to perform the semantic functions of a discipline. In this light, again, a 'function first' rather than 'form first' approach is ideal, starting with

the purpose of particular noun phrase within the text, before going on to attend to lexis then form.

5.4.4 The EGAP vs. ESAP debate

The findings from the current study are relevant to the ongoing debate in EAP regarding the level of disciplinary specificity to be aimed for within the classroom, whether the best model is English for *General* Academic Purposes or English for *Specific* Academic purposes. On one side of this argument are those who argue that EAP must ‘involve teaching literacy skills which are appropriate to the purposes and understandings of particular academic and professional communities’ (Hyland, 2002b, p. 385) and that the optimum model for the teaching of academic discourse is an embedded ‘discipline and context specific’ one (Tribble and Wingate, 2012, p. 492). On the other, are those, often coming from a more pragmatic perspective, who argue that in most practical circumstances an EGAP approach is more suitable and that it is the students themselves who should take the responsibility to learn about the conventions and expectation of their subjects (e.g. de Chazal, 2013). A middle ground between these perspectives is provided by Bruce (2011, p. 7) who argues that ‘the challenge facing the EAP course designer and teacher is to equip students with the tools to deconstruct and make sense of the texts and related discourses of their discipline’.

The findings of the current study in that they reveal how language constitutes academic, and indeed disciplinary, purpose, fall on the ESAP side of the debate in that the phraseological patterning of grammatical keywords can be linked to discipline, either at the level of meta-disciplinary features of student Humanities and Social Sciences writing (Durrant, 2015) or at the level of individual disciplines. However, whether an EGAP or ESAP approach should or can be taken often depends on practical circumstance. Embedded subject-specific in-session classes afford the opportunity to attend in detail to how language and discourse link to

specific disciplinary communicative purpose. In contrast, on pre-sessional courses in which students cannot easily be streamed by discipline, the focus, to an extent, has to be on language, discourse and skills that have a 'transferable' quality.

Whilst the idea of a 'generic' academic language or 'generic' academic skills runs counter to the accumulated evidence from research into disciplinary discourse in both professional and student academic writing, an argument can possibly be made for the teaching of 'meta-disciplinary' literacy skills, shared linguistic purposes and characteristics within wider groupings of academic subjects. The findings of the current study cohere with those of Durrant (2015) in suggesting there are clusters of disciplines which share particular purposes and related discoursal characteristics. Indeed, even the findings of variation in the current study entailed greater or lesser frequency of use of otherwise shared linguistic features.

Greater awareness of which disciplines cluster and what particular characteristics they share can inform a 'middle path' between EGAP and ESAP which allows for a degree of 'meta-disciplinary' differentiation when, for practical purposes, finer disciplinary differentiation is not possible. Attending more to 'meta-disciplinary' literacy skills within EGAP curriculums is arguably an effective way of, as Bruce (2011) calls for, better equipping students with the tools to understand the expectations of their degree programmes. For this purpose, more research exploration of continuities, the shared characteristics of student writing in disciplines which cluster or overlap, would be of particular pedagogical value.

5.5 Reflections on the Methodology

In this project I have adopted with a few minor adaptations the approach towards comparison and contrast of disciplinary discourses which Groom (2007) pioneered in his study of professional academic writing in History and Literature, that of identifying salient grammatical words in research corpora and then conducting an inductive phraseological analysis of the semantic patterning of these grammatical

keywords. Within this process I have employed the concepts of 'semantic sequences' (Groom, 2007; Hunston, 2008) and 'semantic motif' (Groom, 2007). Here, it is useful to be reminded of the exact meaning of each of these terms:

Semantic sequences are recurring sequences of words or phrases that may be very diverse in form and which are therefore usually characterised as sequences of meaning elements rather than formal sequences. (Hunston, 2008, p. 171)

A semantic motif is a group of semantic sequences or other phraseological items which can be grouped together insofar as they share a similar broad meaning ... As the analysis progressed, I began to notice that many of the individual semantic sequences that I had identified seemed to express similar meanings in slightly different ways, and could thus be seen as falling into larger groups of semantically related items. (Groom, 2007, p. 102).

This section will consider, in light of the current study, both the efficacy and limitations of the notions of semantic sequence and semantic motif, as well as the strengths and weaknesses of the methodology as a whole.

5.5.1 Strengths of the methodology

Both Hunston (2008) and Groom (2007) argue for the efficacy of an inductive semantically-focused analysis of phraseological patterning of grammatical keywords, arguing this is *particularly* useful as an approach for revealing the features of disciplinary discourses. Groom argues (2007, p. 282) 'such approaches allow us to address in a very direct and open-ended way the key question of how abstract epistemological constructs ... manifest themselves in conventionalised linguistic forms'. Hunston (2008, p. 292) argues that in terms of a pattern-driven approach 'a targeted search based on 'small words' is 'potentially most useful' because '[t]hese unpromising words ... reveal a surprising amount about the epistemology and ideology of disciplines, because they reveal phraseologies that

are *linked to recurrent meanings and functions rather than to subject-matter*' (my emphasis).

The findings of my study provide further strong evidence to support these assertions by Hunston (2008) and Groom (2007), this time in the context of analysis of student academic writing. The focus on key grammatical words has uncovered a wide range of repeated semantic functions expressed by phraseologies within the student texts that, as has been shown in section 5.2 of this Chapter, have strong and wide-ranging links to disciplinary purposes within the texts, both those shared across student writing in the Social Sciences and Humanities as well as those more particular to specific 'soft' disciplines. Moreover, these phraseologies represent a very large part of the content of the student texts: the five target words alone make up an average of 10% of the text and with their phraseologies represent well over 50% of the text. Thus, the use of grammatical rather than content words as a starting point does indeed very effectively enable the researcher access to *recurrent meanings and functions* that are disciplinary in nature.

The findings of the current study demonstrate that, as an approach, this methodology proves to be potentially an extremely useful one for 'fleshing out' the limited knowledge at a lexico-grammatical level we have thus far accrued about disciplinary features of student texts. As has been shown in section 5.2.1 above, the phraseological findings of the current study show how purposes particular to student Humanities and Social Sciences writing outlined by Durrant (2015) are realised repeatedly within and across texts within this disciplinary grouping. However, in contrast to Durrant's focus on disciplinary lexical bundles, the phraseological approach centred on grammatical keywords used in the current study reveals linguistic features which have a much greater level of frequency within particular actual instances of student writing. As shown in section 5.4.1, when a single text was investigated, only two of Durrant's (2015) Humanities and Social Sciences bundles were found within it; in contrast, phraseologies identified in the current study were found in all but three sentences of the text, and very often occurred twice or more within sentences.

Furthermore, it can be argued that an inductive methodology such as used in the current study which starts with the most salient grammatical words is superior to intuitive approaches as a device for getting at important repeated features of the discourse: Nesi and Gardner (2012) conducted a targeted search for IF-THEN sequences within essays in BAWE which, although revealing interesting disciplinary contrasts with regards to argumentation, because of its top-down intuitive nature, did not necessarily reach linguistic features which occur frequently within the target texts: no examples of IF-THEN sequences were found in the analysis of the essay text in described in section 5.4.1.

Another advantage which could be argued, possibly a little more tentatively, for the approach adopted within the current study is the ‘wide-lens’ view of the target textual data it potentially affords. This argument can be illustrated using the target keyword **that**. Many studies have dealt with the use of *that* in relation to other particular linguistic features. There have been a range of studies involving ‘*it ... that*’ sequences (e.g. Hewings and Hewings 2002, Groom 2005, Thompson 2009) and studies looking at *that* as it behaves with particular verb or noun forms (e.g. Charles 2006 and 2007). Each of these studies contributes to our understanding of the use of *that* in phraseologies for purposes such as creating stance, reporting arguments and presenting evidence. However, a completely bottom-up approach to phraseological analysis of *that* could be argued to allow for a wider and more pedagogically useful perspective on the range of ‘voice’ usages *that* makes possible. Within this wider view, ‘*It ... that*’ sequences, for example, represent a subset of ‘voice’ phraseologies. Again, we can see this in the analysis of a single student text described in section 5.4.1. Only one ‘*It ... that*’ sequence occurred within the text, although there were a large number ‘voice’ phraseologies using *that* to be found which are now listed here: *ideas that ...; Marxist and Anarchist ideologies of resistance that; I shall make the case that ...; They openly declare that ...; the conviction that ...; Foucault means to say that ...; Marxism and Anarchism both hold that ...; Newman’s conclusion is that ...; and, it can be seen that*

The concepts of ‘semantic sequence’ and ‘semantic motif’, with their focus on ‘semantic similarity but formal variation’ (Hunston, 2008, p. 292) are crucial in this bottom-up phraseological method to the affording of a wide-lens view of groupings of semantic functions and the range of ways in which these can be realised linguistically within texts. Within the findings of the current study, the types of phraseologies revealed can be argued to sit at various points on a cline from structurally ‘tighter’ semantic sequences, such as the type typical in *of* phraseologies, to ‘looser’ phraseologies which fall within the concept of ‘motif’ but stretch the concept of ‘semantic sequence’ such as the ‘commentary’ phraseologies for *this*. In analysis and categorisation of *this* phraseologies for commentary I have possibly taken liberties, considering their very wide variability, in seeing the phrases following *this* as ‘semantic elements’, but would argue that the ability to use the idea of ‘semantic sequence’ flexibly is advantageous in operationalising this method as usefully as possible.

5.5.3 Limitations of the methodology

Whilst overall I would make a very positive argument with regards to the usefulness of Groom’s (2007) methodology, there are some limitations which need also to be acknowledged. One such limitation is one that Groom (2007, p. 277) discusses within his own thesis: a focus on grammatical as opposed to lexical words in the keyword analyses yields more data than can be realistically dealt with, especially considering the time-consuming nature of the bottom-up analytical procedure in the follow-up analysis of each item. In Groom’s study of the 55 grammatical keywords that were generated, he was only able to closely analyse 19. This involved for each item the analysis of a 100-line concordance sample for two disciplinary sub-corpora, so, in all, attention to 3,800 concordance lines plus the extra two samples for each discipline he analysed for *of* taking this to a total of 4,200 concordance lines analysed within Groom’s project. Due to, firstly, the smaller size of my corpus meaning that it was decided that larger random samples were needed

to better ensure representativeness, secondly, my use of a pooled random sample to generate semantic categories in the first instance, and, thirdly, the institutional dimension of the research design entailing four disciplinary/institutional sub-corpora, the analysis of only five target grammatical keywords in my study involved, in analysis of the first samples alone, attention to five 300-line samples for each word (the pooled sample plus the four sub-corpora samples) meaning a total of 6,000 concordance lines. However, the limited number of items that I have been able to closely investigate in a qualitative way in my project does not necessarily represent a serious limitation of the study considering the text coverage the phraseologies around these five words involve. Indeed, an argument could be made for an even narrower and more fine-grained focus involving one or two keywords, especially those that have a high percentage frequency within the target texts.

Whilst affording extremely good text coverage and thus leading to a very detailed understanding of 'recurrent meanings and functions' (Hunston 2008, p. 292) within the target discourses, another limitation to this approach is the fact that, whilst it reveals repeated phraseological features of the discourses, it does not on its own provide information about where these revealed phraseologies are located within texts and how they relate to each other within the context of texts. The process of highlighting the phraseologies within one particular text, as described in section 5.4.1, provided a glimpse of how the phraseologies are situated in relation to each other and the whole within the context of a single text. A fruitful future use of the methodology might be in combination with analyses of whole texts to get a sense of how the phraseologies typically work together to constitute an instantiation of the target genre.

A further methodological limitation, also discussed by Groom (2007) within his thesis, is the fact that the semantic categories that arise are 'highly dependent on the knowledge and intuitions of the human analyst' (Groom, 2007, p. 276). The semantic categories I have identified within my analysis are of an interpretive nature: a different analyst may come up with categories partly or even wholly

different from my own. However, that said, when I conducted my phraseological analysis of *as*, I deliberately did this entirely without reference to Groom's analysis of *as* within his thesis and there was still a considerable degree of overlap in the categories produced in both studies. Whilst not denying its overall highly qualitative nature, It could be argued that the fact that two researchers independently arrived at very similar phraseological pictures for the same target keyword within 'soft' disciplines (albeit at different levels) goes some way towards assuaging worries as to a total lack of useful replicability of this approach.

A final limitation which is related only to my particular use of the methodology, and this limitation is one which is wider than the methodology alone, is the size of my sub-corpora and the impact of this on comparability of the sub-corpora and reliability of the results. As has been seen in the discussion of results in Chapter 4, because of the relatively small size of sub-corpora and small number of contributors, individual writer idiosyncrasy has sometimes had a noticeable effect on frequency of phraseologies in particular sub-corpora. This effect has been amplified for the POL/IR-Y sub-corpora which is roughly half the size of other three sub-corpora and involves the fewest individual writer contributors. However, this problem will always be one for researchers wanting to investigate student coursework writing which, as discussed in Chapter 3, is notoriously difficult data to collect in large volumes. It can be argued that the 'control' afforded by the institutional dimension in my research design has helped to improve the validity and consequent reliability of findings.

Despite limitations, overall, a very strong claim can be made for the usefulness of this approach, particularly in studies of student academic writing which aim to inform EAP and academic writing pedagogy. A phraseological analysis which takes as its starting point key grammatical words allows the researcher to get at repeated features of the text which have very high 'operationalization' value for teaching (Bruce, p. 6) in that they represent coverage of a considerable proportion of target texts. A semantic approach to the categorisation of phraseologies uncovers repeated functions and purposes within the target student discourse. These can

then be used as the starting point in an approach to the teaching of discipline-specific academic writing which begins with the highlighting key purposes then moves to focus on how these purposes are realised lexico-grammatically.

Chapter 6: Conclusions

This chapter will draw conclusions regarding the extent to which the objectives set out in the introduction have been achieved within this project. It will also identify limitations and suggest fruitful avenues for further student academic writing research with a phraseological focus.

6.1 Achievements of the thesis

The following objectives for this project were set out at the end of Chapter 1. All objectives can be argued to have been broadly met.

- To reveal the extent to which there are similarities or differences to be found at the lexico-grammatical level in good student History and Politics/International Relations writing at two UK HE institutions.
- To assess how far any similarities or differences which are found can be argued to be accounted for by the influence of ‘discipline’ or by the influence of ‘institution’
- In doing the above, to uncover within the target texts lexico-grammatical features which may have pedagogical value for an EAP teaching context.
- Also, in doing the above, consider the usefulness of Groom’s (2007) approach when it is applied to the analysis of student, as opposed to professional, written academic discourses.

The thesis has achieved the linked objectives of identifying lexico-grammatical similarities and differences in the writing in the two disciplines at two institutions as represented by the research corpora. In the analysis of semantic patterning of five grammatical keywords, *of*, *and*, *that*, *as* and *this*, within the four disciplinary/institutional sub-corpora a large range of similarities and a smaller number of differences were identified. Both similarities and differences could be largely explained with reference to discipline at the level of meta-disciplinary grouping or at the level of individual discipline. Phraseological similarities could be

linked to Durrant's (2015) characteristics of student Humanities and Social Sciences writing. Differences were not absolute but a matter of relative emphasis and could be for the most part linked to differing foci and emphases with the essay questions of the respective disciplines which could be argued to link to a greater emphasis on the Participant and Discourse domains (Gosden, 1993) in Politics/International Relations in contrast to a greater emphasis on the Real-World domain (ibid.) in History. There were no instances in which a pattern of similarity or difference could be explained by particular institutional context, and the results suggest that in terms of their respective influence on the lexico-grammatical features of student writing in the target disciplines at the target institutions, discipline seems to override institution.

It could be ventured that the thesis has been particularly successful with regards to uncovering pedagogically valuable lexico-grammatical features within the target texts. As has been argued in Chapter 5, in contrast to other recent foci of research into student texts, identification of phraseological patterning around grammatical keywords equates to an extremely high level of coverage of the language in any particular example text within the target discourse. This means such an analysis potentially yields a considerable amount of teachable material, or, to quote the same passage from Bruce (2011, p. 6) again, has the potential to 'operationalize the wider phenomenon of academic subject discourses realized in texts'.

With regards to testing the efficacy of Groom's (2007) approach in the context of research into student academic writing, the thesis has also achieved its objectives. In contrast to the very large corpora which are relatively easy to assemble for research into professional academic writing, the size of research corpora that are able to be built to study student writing are much smaller. My corpora were extremely small at approximately 100,000 words each in comparison to Groom's research corpora of between one and four million words, and there were thus concerns about whether the approach would translate effectively. The larger random sample size of 300 lines and the institutional dimension of the research design entailing four sub-corpora for analysis and comparison/contrast of the two

target disciplines were both used to compensate as far as possible for the smaller sub-corpora size.

Overall, it can be argued that Groom's approach has translated successfully to a student writing context. Evidence supporting this claim includes the fact the frequencies for the five target keywords were extremely similar across sub-corpora and that the pattern of frequency of particular semantic patterns were also very often similar across sub-corpora suggesting the corpora were large enough for these consistencies to emerge. This may not have been the case had the focus been on 'content' keywords, but, it can be argued quite confidently that a focus on *grammatical* keywords could serve as an effective analytical starting point for research involving the relatively small corpora it is possible to build in the study student academic texts. A qualification to this argument is the noticeable impact of individual writer idiosyncrasy on frequencies of some patterns. However, the fact that this issue only emerged in a few instances limits the seriousness of this concern.

6.2 Limitations and suggestions for future research

As has been discussed immediately above, a limitation of this study has been the relatively small size of the research corpora, particularly in the case of the smaller POL/IR-Y sub-corpus, and the potential resulting influence of individual writer idiosyncrasy on findings. A related but arguably more significant limitation is the fact that such research corpora as have been the focus of this study can only ever represent a snapshot of student writing within the target disciplines as realised at a particular time in particular institutions by a particular small constellation of individual student writers. However, this is a weakness inherent in all research into student texts, the only solution to which lies in studies of student disciplinary writing being replicated and repeated multiple times in order to build a more longitudinal picture of repeated patterns and characteristics. A further limitation has been the fact that within this study the texts have been treated solely as bodies

representing a particular discipline at a particular institution. This may have masked the shared or differentiating characteristics related to the focus of the essay questions being addressed or the sub-genre of the Essay genre family (Nesi and Gardner, 2012) to which each text belonged.

It has been argued above that grammatical keywords have proven a useful starting point for phraseological analysis of student academic texts, and there are a wide range of directions in which future research related to that of the present study could be pursued. Further corpus-driven studies taking this approach could be usefully carried out with a focus on different disciplinary groupings as well as on other genres of student text. This approach could be used to compare the phraseological features of different genres within the same discipline, or, in response to the limitation outlined above, be used in a comparative analysis of Essay genre family sub-genres either within or across disciplinary groupings.

This study has sought to comprehensively categorise the semantic patterns of five grammatical keywords resulting in considerable depth in the analysis of the two disciplines involved. However, considering the time-consuming nature of the procedure, an alternative approach might be to look at one important grammatical keyword across a wider range of disciplines. ‘The voice of the essay writer’ and ‘the voice of others’ patterns for **that**, for example, might prove a particularly rich vein for analysis in this respect affording as it would cross-disciplinary insights into the way ideas are typically framed and voices within texts portrayed.

Finally, as already discussed in the last section of Chapter 5, whilst a strength of the approach is its ability to reveal recurring patterns of meaning across the texts of a discourse, a weakness is that on its own it provides the researcher with almost no understanding of how these patterns behave and interact within the context of individual texts. Therefore another useful direction that future research may take might involve the triangulation of phraseological analysis of grammatical keywords with an analysis of how they manifest within whole instances of texts in the target discourse.

6.3 Final thoughts

I embarked on this project in search of answers to EAP teaching motivated questions, related particularly to the relative value of different types of focus in the teaching of the Essay genre in academic writing. It was my position then that the wrong aspects of essay writing were being privileged in much EAP teaching practice and that students were thus being left with a highly refined knowledge of the structural 'shell' of an academic essay, but with woefully limited knowledge of the important purposes and functions within such texts and how these were operationalized at the lexico-grammatical level. In five subsequent years of experience and experimentation as an EAP practitioner my position that more attention needs to be paid to lexico-grammatical features of target academic discourses has only been further strengthened.

Turner's (2004) critique of the prioritising of macro-structural elements and of fluency over attention to word and phrase level features and to accuracy, and her related call for a reconceptualization of EAP practices to centre on 'language as constitutive' of academic purpose is now twelve years old. However, it is my experience that there is a considerable lag in practice and that the 'skills' mindset is still widely ingrained and strong.

It is hoped that the research that been carried out in this project contributes usefully to a the growing body of student academic writing research which is both raising awareness of and building understanding of the distinctive features of student academic discourse, and that this will in turn feed into more language and discourse-sensitive EAP writing pedagogies.

References

- Allen, D. 2009. Lexical bundles in learner writing: An analysis of formulaic language in the ALESS Learner Corpus. *Komaba Journal of English Education* 1. 105-127.
- Alsop, S. & Nesi, H. 2009. Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora* 4(1). 71-83.
- Askehave, I. & Swales, J. 2001. Genre Identification and Communicative Purpose: A Problem and Possible Solution. *Applied Linguistics* 22(2). 195-212.
- Baker, P. 2004. Querying keywords: questions of difference, frequency and sense in keywords analysis. *Journal of English Linguistics* 32(4). 346-359.
- Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
- Basturkmen, H. 2012. A genre-based investigation of discussion sections in research articles in Dentistry and disciplinary variation. *Journal of English for Academic Purposes* 11. 134-144.
- Bawarshi, A. & Reiff, M.J. 2010. *Genre: An introduction to history, theory, research, and pedagogy*. West Lafayette: Parlor Press.
- Bawarshi, A. & Reiff, M. 2011. Tracing Discursive Resources: How Students Use Prior Genre Knowledge to Negotiate New Writing Contexts in First-Year Composition. *Written Communication* 28(3). 312-337.
- Bazerman, C. 1981. What written knowledge does: three examples of academic discourse. *Philosophy of the Social Sciences* 11. 361-387.
- Bazerman, C. 1994. Systems of Genres and the enactment of Social Intentions. In Freedman, A. and Medway, P. (eds.), *Genre and the New Rhetoric*, 79-101 London: Taylor and Francis.
- Becher, T. 1981. Towards a definition of disciplinary cultures. *Studies in Higher Education* 6(2). 109-122.
- Becher, T. 1987a The disciplinary shaping of the profession. In Clark, B.R. (ed.), *The Academic Profession*. Berkley: University of California Press.
- Becher, T. 1987b. Disciplinary discourse. *Studies in Higher Education* 12(3). 261-274.
- Becher, T. 1989. *Academic Tribes and Territories*. Buckingham: Open University Press.
- Becher, T. 1994. The significance of disciplinary differences. *Studies in Higher Education* 19(2). 151-161.

Becher, T. & Trowler, P.R. 2001. *Academic Tribes and Territories: Second Edition*. Buckingham: Open University Press.

Berkenkotter, C. & Huckin, T. 1993. Rethinking Genre from a Sociocognitive perspective. *Written Communication* 10(4). 475-509.

Berkenkotter, C. & Huckin, T. 1995. *Genre knowledge in disciplinary communication: Cognition/culture/power*. Hillsdale, New Jersey: Lawrence Erlbaum.

Biber, D. 1989. A typology of English texts. *Linguistics* 27. 3-43.

Biber, D., Gray, B. & Ponpoon, K. 2011. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly* 45(1). 5-35.

Bondi, M. & Silver, M. 2004. Textual Voices: A Cross Disciplinary Study of Attribution in Academic Discourse. In Anderson, L. & Bamford, J. (eds.). *Evaluation in Oral and Written Discourse*, 117-136. Rome: Officina Edizioni.

Bondi, M. 2005. Metadiscursive Practices in Academic Discourse: variation Across Genres and Disciplines. In Bamford, J. and Bondi, M. (eds.) *Dialogue within Discourse Communities*, 3-30. Tuburgen: Niemeyer.

Bondi, M. 2006. 'A case in point': Signals of narrative Development. In Hyland, K. & Bondi, M. (eds.) *Academic Discourse Across disciplines*, 49-74. Bern: Peter Lang.

Borg, E. 2003. Discourse Community. *ELT Journal* 57(4). 398-400.

Bowker, L. & Pearson, J. 2002. *Working with specialised language: A Practical guide to using corpora*. London: Routledge.

Braine, G. 1989. Writing in science and technology: an analysis of assignments from ten undergraduate courses. *English for Specific Purposes* 8. 3-15.

Braun, V. & Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2). 77-101.

Breeze, R. 2011. Disciplinary values in legal discourse: a corpus study. *Iberica* 21. 93-116.

Bruce, I. 2008. *Academic Writing and Genre: A Systematic Analysis*. London: Continuum.

Bruce, I. 2009. Results sections in sociology and organic chemistry articles: A genre analysis. *English for Specific Purposes* 28(2). 105-124.

Bruce, I. 2010. Textual and discoursal resources used in the essay genre in sociology and English. *Journal of English for Academic Purposes* 9. 153-166.

Bruce, I. 2011. *Theory and Concepts of English for Academic Purposes*. Basingstoke: Palgrave Macmillan.

Bruce, I. 2013. A role for genre-based pedagogy in academic writing instruction?: an EAP perspective. *Text Special Issue* 21, October 2013. Retrieved from: <http://www.textjournal.com.au/speciss/issue21/content.htm>

Candlin, C. & Plum, G. 1999. Engaging with challenges of discursivity in academic writing. Researchers, students and tutors. In Candlin, C. & Hyland, K. (eds.), *Writing: Texts, processes and practices*, 193-218. London: Longman.

Carter, M. 2007. Ways of knowing, Doing and Writing in the Disciplines. *College Composition and Communication* 58(3). 385-417.

Chan, S-H. & Tan, H. 2010. Extracting and comparing the intricacies of metadiscourse of two written persuasive corpora. *International Journal of Education and Development using Information and Communication Technology* 6(3). 124-146.

Charles, M. 2006a. The construction of stance in reporting clauses: a cross-disciplinary study of these. *Applied Linguistics* 27(3). 492-518.

Charles, M. 2006b. Phraseological patterns in reporting clauses used in citation: A corpus-based study of these in two disciplines. *English for Specific Purposes* 25(3). 310-331.

Charles, M. 2007. Argument or evidence? Disciplinary variation in the use of the NOUN *that* pattern in stance construction. *English for Specific Purposes* 26(2). 203-218.

Chen, S-H. & Baker, P. 2010. Lexical Bundles in L1 and L2 Academic Writing. *Language Learning and Technology* 14(2). 30-49.

Connor, U. & Upton, T. A. 2004. Preface. In Connor, U. & Upton, T. A. (eds.), *Applied corpus linguistics: A multidimensional perspective* III-V. Amsterdam: Rodopi.

Conrad, S. 2001. Variation among disciplinary texts: a comparison of textbooks and journal articles in biology and history. In Conrad, S. & Biber, D. (eds.), *Variation in English: multidimensional studies*, 94-107. Harlow: Longman.

Conrad, S. 2002. Corpus Linguistics Approaches for Discourse Analysis. *Annual Review of Applied Linguistics* 22. 75-95

Cortes, V. 2004. Lexical bundles in published and student writing in history and biology. *English for Specific Purposes* 23(4). 397-423

Crombie, W.H. 1985. *Process and relation in discourse and language learning*. Oxford: Oxford University Press.

De Chazal, E. 2013. The general-specific debate in EAP: which case is the most convincing for most contexts? *Journal of Second Language Teaching and Research* 2(1). 135-148.

Devitt, A. 1993. Generalising about Genre: New Conceptions of an Old Concept. *College Composition and Communication* 44(4). 573-86.

Donahue, J. 2012. Using systemic functional linguistics in academic writing development: An example from film studies. *Journal of English for Academic Purposes* 11. 4-16.

Durrant, P. & Matthews-Aydinli, J. 2011. A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes* 30(1). 58-72.

Durrant, P. 2014. Discipline and Level Specificity in University Students' Written Vocabulary. *Applied Linguistics* 35(3). 328-356.

Durrant, P. Advance Access publish April 19, 2015. Lexical Bundles and Disciplinary Variation in University Students' Writing: Mapping the Territories. *Applied Linguistics*. 1-30.

Ebeling, S.O. 2011. Recurrent word combinations in English student essays. *Nordic Journal of English Studies* 10(1). 49-76.

Firth, J.R. 1968. A synopsis of linguistic theory 1930-55, in F.R. Palmer (ed.) *Selected Papers of J.R. Firth 1952-59*, 168-205. London: Longman.

Flowerdew, L. 2005. An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies. *English for Specific Purpose* 24(3). 321-332.

Francis, G. 1986. Anaphoric nouns. *Discourse Analysis Monographs*, No 11. University of Birmingham: English Language Research.

Francis, G. 1994. Labelling discourse: an aspect of nominal group lexical cohesion. In Coulthard, M. (ed.), *Advances in written text analysis*, 83-101. London: Routledge.

Francis, G., Manning, E. & Hunston, S. 1998. *Collins COBUILD grammar patterns 2: nouns and adjectives*. London: HarperCollins.

- Gardner, S. 2008. Mapping ideational meaning in a corpus of student writing. In C Jones and E. Ventola (eds.), *New Developments in the Study of Ideational Meaning: From Language to Multimodality*, 169-188. London: Equinox Publishing.
- Gardner, S. 2012. Genres and registers of student writing: an SFL perspective on texts and practices. *Journal of English for Academic Purposes* 11(1). 52-63.
- Gee, P. 1989. What Is Literacy? *Journal of Education* 171(1). 18-25.
- Gimenez, J. 2012. Disciplinary epistemologies, generic attributes and undergraduate academic writing in nursing and midwifery. *Higher Education* 63. 401-419.
- Gledhill, C. 2000. The discourse function of collocation in research article introductions. *English for Specific purposes* 19(2). 115-135.
- Gosden, H. 1993. Discourse Functions of a Subject in Scientific Research Articles. *Applied Linguistics* 14(1). 56-75.
- Gries, S.T. 2010. Corpus linguistics and theoretical linguistics: A love-hate relationship? Not necessarily. *Special Issue of international Journal of Corpus Linguistics* 15(3). 327-343.
- Groom, N. 2005. Pattern and meaning across genres and disciplines: an exploratory study. *Journal of English for Academic Purposes* 4(3). 257-277.
- Groom, N. 2007. Phraseology and epistemology in humanities writing: a corpus-driven study. Unpublished doctoral dissertation, University of Birmingham, Birmingham, UK.
- Groom, N. 2010. 'Closed-class keywords and corpus-driven discourse analysis.' In Bondi, M. & Scott, M. (eds.), *Keyness in Texts*, 59-78. Amsterdam and Philadelphia: Benjamins.
- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B. & Kantor, R. 1996. *A study of writing tasks assigned in academic degree programs*. Princeton, NJ: Educational Testing Service.
- Halliday, M.A.K. 1978. *Language as Social Semiotic*. London: Edward Arnold.
- Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. 2nd Ed. London: Edward Arnold.
- Hardy, J.A. & Römer, U. 2013. Revealing disciplinary variation in student writing: a multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora* 8(2). 183-207.

Henderson, A. & Barr, R. 2010. Comparing indicators of authorial stance in psychology: students' writing and published research articles. *Journal of Writing Research* 2(2). 245-264.

Hewings, M. & Hewings, A. 2002. "It is interesting to note that ...": a comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes* 21(4). 367-383.

Hoey, M. 1983. *On the surface of discourse*. London: George Allen & Unwin.

Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Hsieh, H. & Shannon, S. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15(9). 1277-1288.

Hunston, S. & Francis, G. 1999. *Pattern Grammar*. Amsterdam: John Benjamins.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, S. 2008. Starting with the small words. *International Journal of Corpus Linguistics* 13(3) 272-295.

Hyland, K. 2000. *Disciplinary discourses: Social interactions in academic writing*, Harlow: Pearson Education.

Hyland, K. 2001. Humble servants of the discipline? Self-mention in research articles, *English for Specific Purposes* 20(3). 207-226.

Hyland, K. 2002a. Authority and invisibility: Authorial identity in academic writing, *Journal of Pragmatics* 34. 1091-1112.

Hyland, K. 2002b. Specificity revisited: how far should we go now? *English for Specific Purposes* 21(4). 385-395.

Hyland, K. 2005. Stance and engagement: a model of interaction and academic discourse. *Discourse Studies* 7(2). 173-192.

Hyland, K. 2007. Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing* 16. 148-164.

Hyland, K. 2008. Genre and academic writing in the disciplines. *Language Teacher* 41(4). 543-562.

Hyland, K. 2009. Writing in the disciplines: Research evidence for specificity. *Taiwan International ESP Journal* 1(1). 5-22.

- Hyon, S. 1996. Genre in 3 traditions for ESL. *TESOL Quarterly* 30(4). 693-772.
- Ivanic, R. 1998. *Writing and identity: the discursal construction of identity in academic writing*. Amsterdam: John Benjamins.
- Johns, A.M. (ed.) 2002. *Genre in the classroom: Multiple perspectives*, Mahwah, New Jersey: Lawrence Erlbaum.
- Johns, A.M., Bawarshi, A., Coe, R., Hyland, K., Paltridge, B., Reiff, M.J. & Tardy, C. 2006. Crossing the boundaries of genre studies: Commentaries by experts. *Journal of Second Language Writing* 15. 234-249.
- Johns, A.M. (2008). Genre awareness for the novice academic students: An ongoing quest. *Language Teaching* 41(2). 237-249.
- Johnson, S. & Ensslin, A. 2006. Language in the news: Some reflections on Keyword Analysis Using Wordsmith Tools and the BNC. *Leeds Working Papers in Linguistics and Phonetics*, 11.
- Kroll, B. 1979. A survey of the writing needs of foreign and American college freshmen. *English language Teaching Journal* 23. 219-227.
- Lave, J. & Wenger, E. 1991. *Situated Learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Lea, M.R. & Street, B.V. 1998. Student writing in higher education: an academic literacies approach. *Studies in Higher Education* 23(2). 157-172.
- Lea, M.R. & Street, B.V. 1999. Writing as academic literacies: understanding textual practices in higher education. In Candlin, C.N. & Hyland, K. (eds.), *Writing: Texts, Processes and Practices*, 62-81. London: Longman.
- Lee, D. & Chen, X. 2009. Making a bigger deal of smaller words: function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing* 18(4). 281-296.
- Leech, G. 1992. Corpora and theories of linguistic performance. In Svartvik, J. (ed.), *Directions in corpus linguistics: Proceedings of the Nobel Symposium 82*, 105-122. New York: Mouton de Gruyter.
- Li, T. & Wharton, S. 2009. Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal of English for Academic Purposes* 11. 345-356.
- Lillis, T. 1997. New voices in academia? The regulative nature of academic writing conventions. *Language and Education* 11(3). 182-199.

- Louw, W.E. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Baker, M., Francis, G. & Tognini-Bonelli, E. (eds.) *Text and Technology: In Honour of John Sinclair*, 157-76. Amsterdam: John Benjamins.
- Louw, W.E. 2010. "The painting is where the paint is not": Reflections on the "Bootcamp Debate". *Special Issue of international Journal of Corpus Linguistics* 15(3). 344-353.
- Mahlberg, M. 2006. Lexical cohesion: corpus linguistic theory and its application in English language teaching. *International Journal of Corpus Linguistics* 11(3). 363-383.
- Malavasi, D. & Mazzi, D. 2010. History vs. marketing: Keywords as a clue to disciplinary epistemology. In Bondi, M. & Scott, M. (eds.), *Keyness in Texts*, 169-184. Amsterdam: John Benjamins Publishing Company.
- Martin, J.R. 1984. Language, Register and Genre. In Christie, F. (ed.), *Language Studies: Children's Writing: Reader*, 21-9. Geelong: Deakin University Press. Reprinted with revisions in Burns, A. & Coffin, C. (eds.), 2001. *Analyzing English in a Global Context*, 149-66. London: Routledge.
- Martin, J.R. 1993. In M.A.K Halliday and J.R. Martin, *Writing Science: Literacy and Discursive Power*. London: The Falmer Press.
- Martin, J.R. 1997. Analysing genres: functional parameters. In Christie, F. & Martin (eds.), *Genre and institutions*, 3-39. London: Cassell.
- Martin, J.R. 2009. Genre and language learning: A social semiotic perspective. *Linguistics and Education* 20. 10-21.
- McCarthy, M. 1994. It, this and that. In Coulthard, M. (ed.), *Advances in written discourse analysis*, 266-275. London: Routledge.
- McEnery, T. & Wilson. A. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T., Xiao, R. & Tono, Y. 2005. *Corpus-Based language Studies: an advanced resource book*. Oxon: Routledge.
- McEnery, T. & Gabrielatos, C. 2006. English corpus linguistics. In B. Aarts & A. McMahon (eds.), *The Handbook of English linguistics*, 33-71. Oxford: Blackwell.
- McEnery, T. & Hardie, A. 2010. On two traditions in corpus linguistics, and what they have in common. *Special Issue of international Journal of Corpus Linguistics*, 15(3). 384-394.

McEnery, T. & Hardie, A. 2012. *Corpus Linguistics*. Cambridge: Cambridge University Press.

Meyer, C.F. 2002. *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.

Miller, C.R. 1984. Genre as social action. *Quarterly Journal of Speech* 70. 151-167.

Miller, C. & Shepherd, D. 2004. Blogging as Social Action: A Genre Analysis of the Weblog. In Gurak, L.J., Antonijevic, S., Johnson, L., Ratliff, C. and Reyman, J. (eds.), *Into the Blogosphere: rhetoric, Community and Culture of Weblogs*. Retrieved from: http://blog.lib.umn.edu/blogosphere/blogging_as_social_action_a_genre_analysis_of_the_weblog.html

Moore, T. & Morton, J. 1999. *Authenticity in the IELTS Academic module Writing Test: A comparative study of Task 2 items and university assignments*. IELTS Research Reports 2, 64-106. Canberra: IELTS Australia.

Moore, T. & Morton, J. 2005. Dimensions of difference: a comparison of university writing and IELTS writing. *Journal of English for Academic Purposes* 4(1). 43-66.

Mukherjee, J. 2010. Corpus linguistics versus corpus dogmatism. *Special Issue of international Journal of Corpus Linguistics* 15(3). 370-378.

Nesi, H. 2008. BAWE: an introduction to a new resource. In Frankenberg-Garcia, A. Rkibi, T., Braga da Cruz, M., Carvalho, R., Direito, C. & Santos-Rosa, D. (eds.), *Proceedings of the Eight teaching and Language Corpora Conference*, 239-46. Lisbon, Portugal: ISLA.

Nesi, H, Sharpling, G. & Ganobcsik-Williams, L. 2004. The design, development and purpose of a corpus of British student writing. *Computers and Composition* 21(4). 439-450.

Nesi, H. & Gardner, S. 2012. *Genre across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.

Newmann, R., Parry, S. & Becher, T. 2002. Teaching and learning in their disciplinary contexts: A conceptual analysis. *Studies in Higher Education* 27(4). 405-417.

Nishina, Y. 2010. Evaluative Meanings and Disciplinary Values: A Corpus-based Study of Adjective Patterns in research Articles in Applied Linguistics and Business Studies. Unpublished doctoral dissertation, University of Birmingham, Birmingham, UK.

North, S. 2005. Disciplinary variation in the Use of Theme in Undergraduate Essays. *Applied Linguistics* 26(3). 431-452.

Paltridge, B. 1994. Genre Analysis and the Identification of Textual Boundaries. *Applied Linguistics* 15(3). 288-299.

Paltridge, B. 1995a. Working with genre: A pragmatic Perspective. *Journal of Pragmatics* 24. 393-406.

Paltridge, B. 1995b. Analyzing Genre: A Relational Perspective. *System* 23(4). 503-511.

Paltridge, B. 1996. Genre, text type, and the language learning classroom. *ELT Journal* 50(3). 237-243.

Paltridge, B. 2002. Genre, text type, and the English for Academic Purposes (EAP) classroom. In Johns, A. (ed.), *Genre in the classroom: Multiple perspectives*, 73-90. Mahwah, NJ: Lawrence Erlbaum.

Parkinson, J. & Musgrave, J. 2014. Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes* 14. 48-59.

Prior, P. 1994. Response, revision, Disciplinary: A Microhistory of a Dissertation Prospectus in Sociology. *Written Communication* 11(4). 483-533.

Prior, P. 1997. Literate Activity and Disciplinary: The Heterogeneous (Re)production of American Studies Around a Graduate Seminar. *Mind, Culture and Activity* 4(4). 275-295.

Rose D., Rose M., Farrington, S. & Page, S. 2008. Scaffolding academic literacy with indigenous health sciences students: An evaluative study. *Journal of English for Academic Purposes* 7(3). 165-79.

Russell, D.R. 1997. Rethinking Genre in School and Society: An Activity theory Analysis. *Written Communication* 14(4). 504-544.

Russell, D.R. 1997. Rethinking Genre in School and Society: An Activity theory Analysis. *Written Communication* 14(4). 504-544.

Samraj, B. 2002. Introductions in research articles: variations across disciplines. *English for Specific Purposes* 21. 1-17.

Samraj, B. 2008. A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes* 7. 55-67.

Schleppergrell, M. 2004a. Technical writing in a second language: The role of grammatical metaphor. In Ravelli, L. & Ellis, R. (eds.), *Analysing academic writing: Contextualised frameworks*, 172-189. London: Continuum.

Schleppergrell, M. 2004b. *The language of schooling: A functional Linguistics perspective*. Marwah, NJ: Lawrence Erlbaum.

Scott, M. 2001. Comparing corpora and identifying key words, collocations and frequency distributions through the *Wordsmith Tools* suite of computer software. In Ghadessy, M., Henry, A. & Roseberry, R. (eds.), *Small Corpus Studies and ELT*, 47-67. Amsterdam: John Benjamins.

Scott, M. 2010. Problems in investigating keyness, or clearing the undergrowth and marking out trails. In Bondi, M. & Scott, M. (eds.), *Keyness in Texts*, 43-57. Amsterdam: John Benjamins.

Scott, M. & Tribble, C. 2006. *Textual Patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.

Shryer, C.F. 1993. Records as genre. *Written Communication* 10. 200-234.

Shryer, C.F. 1994. The Lab vs. The Clinic: Sites of Competing Genres. In Freedman, A. & Medway, P. (eds.), *Genre and the New Rhetoric*. 105-124. London: Taylor and Francis.

Silver, M. 2006. Introducing Abstract Reasoning: World of Reference and Writer Argument Across Disciplines. In Hyland, K. & Bondi, M. (eds.), *Academic Discourse Across disciplines*, 75-101. Bern: Peter Lang.

Sinclair, J.M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J.M. 2004. *Trust the text: Language, corpus and discourse*. R. Carter (ed.) London: Routledge.

Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Stubbs, M. 1993. British traditions in text analysis: From Firth to Sinclair. In Baker, M., Francis, F. & Tognin-Bonelli, E. (eds.). *Text and technology: In honour of John Sinclair*, 1-36. Amsterdam: John Benjamins.

Swales, J. 1990. *Genre Analysis: English in academic and research settings*, Cambridge: Cambridge University Press.

Swales, J. 2004. *Research Genres: Exploration and Applications*, Cambridge: Cambridge University Press.

Swales, J. 2009. Worlds of Genre – Metaphors of Genre. in Bazerman, C., Bonini, A., & Figueiredo, D. (eds.) *Genre in a Changing World*. Fort Collins, Colorado: The

WAC Clearinghouse and Parlor Press. Retrieved from:
<http://wac.colostate.edu/books/genre/>

Swales, J. 2011. Coda: Reflections on the future of genre and L2 writing. *Journal of Second Language Writing* 20. 83-85.

Swales, J. & Feak, C. 1994. *Academic Writing for Graduate Students: Essential tasks and Skills*. Ann Arbor: University of Michigan Press.

Tardy, C.M. 2011. The history and future of genre in second language writing. *Journal of Second Language Writing* 20(1). 1-5.

Taylor, C. 2008. What is *corpus linguistics*? What the data says. *ICAME Journal* 32. 179-200.

Taylor, C. 2013. Searching for similarity using corpus-assisted discourse studies. *Corpora* 8(1). 81-113.

Teubert, W. 2005. My version of corpus linguistics. *International Journal of Corpus Linguistics* 10(1). 1-13.

Teubert, W. 2009. Corpus Linguistics: An Alternative, *Semen* 27. Retrieved from:
<http://semen.revues.org/8912>

Teubert, W. 2010a. Our brave new world? *Special Issue of international Journal of Corpus Linguistics* 15(3). 354-358.

Teubert, W. 2010b. My brave old world. *Special Issue of international Journal of Corpus Linguistics* 15(3). 395-399.

Thompson, P. 2009. Shared disciplinary norms and individual traits in the writing of British undergraduate. In Gotti, M. (ed.). *Commonality and individuality in Academic Discourse*, 53-82. Bern: Peter Lang.

Tognin-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Tribble, C. 2015. Writing academic English further along the road. What is happening now in EAP writing instruction? *ELT Journal* 69(4). 442-462.

Turner, J. 2004. Language as academic purpose. *Journal of English for Academic Purposes* 3. 95-109.

UCREL, Lancaster University, 1998. *The British National Corpus Sampler Corpus: Explanatory documentation*. Retrieved from:
<http://ucrel.lancs.ac.uk/bnc2sampler/sampler.htm>

Wilks, Y. 2010. Corpus linguistics and computational linguistics. *Special Issue of international Journal of Corpus Linguistics* 15(3). 408-411.

Willis, D. 1990. *The Lexical Syllabus*. London and Glasgow: Collins E.L.T.

Wingate, U. & Tribble, C. 2012. The Best of Both Worlds? Towards an English for Academic Purposes/Academic Literacies Writing Pedagogy. *Studies in Higher Education* 37(4). 481-495.

Woodward-Kron, R. 2004. 'Discourse communities' and 'writing apprenticeship': and investigation of these concepts in undergraduate Education students' writing. *Journal of English for Academic Purposes* 3. 139-161.

Woodward-Kron, R. 2009. "This means that ...": a linguistic perspective of writing and learning in a discipline. *Journal of English for Academic Purposes* 8. 165-179.

Appendix 1: Essay titles

HIST-X

| Essay | Essay title | Topic |
|-------------|---|---|
| HIST3_001_X | What was Goebbels' strategy for conquering Berlin? | Modern Germany |
| HIST3_001_Y | Who brought the wall down, and why? | " |
| HIST3_001_Z | Evaluate the importance of external actors in the formation of the two Germanys | " |
| HIST3_007_P | Compare Smith and Marx's accounts of Original Accumulation | Modern Political Ideas |
| HIST3_007_Q | In Veblen's The Theory of the Working Class which is more important for understanding capitalism, the concept of 'leisure' or 'work'? | " |
| HIST3_009_A | Is the medical activity of medieval women the base of a large iceberg? | Medicine & Society in Medieval Europe |
| HIST3_009_B | Was Islamic medicine merely the systematisation of Greek Medicine? | " |
| HIST3_009_C | 'In the middle ages, magical medicine and religious medicine were essentially different examples of the same thing'. Discuss | " |
| HIST3_009_D | How far did the successes of the Albigensian Crusade owe to the personal achievements of Simon de Monfort? | The Crusades |
| HIST3_009_E | Why was there so much competition between physicians and surgeons in the Middle Ages? | Medicine & Society in Medieval Europe |
| HIST3_009_F | Was the heresy of Waldes an issue of belief or authority prior to 1208? | The Crusades |
| HIST3_010_A | What was the nature of relations between the Islamic world and Byzantium 1000-1300? | The Medieval Islamic World |
| HIST3_010_B | To what extent did the West End of London become a feminised space between 1840 and 1900? | Victorian History |
| HIST3_010_C | For what reasons were the Turks able to assimilate into Islamic society? | The Medieval Islamic World |
| HIST3_010_D | Were the prisons of Victorian London designed more for correction or punishment? | Victorian History |
| HIST3_010_E | Who was the architect of the Muslim triumphs over the crusaders in the late 1180s, Saladin or Nur al-Din? | The Crusades |
| HIST3_010_F | Why did the publication of James Greenwood's "A Night in a Workhouse" in 1866 arouse so much controversy? | Victorian History |
| HIST3_011_R | 'To what extent did Roman institutions and culture survive in early Merovingian Gaul? | Roman History |
| HIST3_011_T | 'What did Julian the apostate seek to achieve in his 18 month rule? What was the legacy of the last Roman Emperor? | Roman History |
| HIST3_011_U | Working-class housing reformers were more interested in the moral than in the material improvement of the London poor. Discuss. | Victorian History |
| HIST3_011_V | 'in what ways did suburban expansion between the 1840s and the 1890s change the nature of the metropolis? | " |
| HIST3_011_W | How effective were efforts at sanitary reform in London between 1850 and 1890? | Victorian History |
| HIST3_012_A | Discuss the interplay between art and academic research stimulated by <i>Copenhagen</i> | The Bomb: Atomic Weaponry and Society in the 20th Century |
| HIST3_012_B | With reference to Oppenheimer, reflect on the moral dilemma of scientists in disseminating research on one hand, and developing weapons of mass destruction on the other. | " |
| HIST3_012_C | Did the medium of print support significant challenges to Victorian societal constructs of gender? | Victorian Social & Political Thought |
| HIST3_012_D | Assess contemporary concerns during the first two decades of the twentieth century over the impact of technology on the future of warfare. | The Bomb: Atomic Weaponry and Society in the 20th Century |
| HIST3_012_E | Why did Matthew Arnold believe that 'culture' was important? | Victorian Social & Political Thought |
| HIST3_013_G | The Cold War between China and America did not really end in 1979.' Discuss. | China and the US during the Cold War |
| HIST3_013_H | Critically evaluate the role of the "China Factor" in Johnson's decision to escalate the war in Vietnam in 1965 | " |
| HIST3_013_I | To what extent can the longevity of the Franco regime be explained in terms of a series of favorable international contexts? | Spain: From Dictatorship to Democracy |
| HIST3_013_J | Was Communist China provoked into entering the Korean War in late October 1950? | China and the US during the Cold War |
| HIST3_013_K | Was Mao to blame for the emergence of the Sino-Soviet split in the mid- and late 1950s? | " |

HIST-Y

| Essay | Essay title | Topic |
|-------|---|--|
| 0019d | Discuss the view of J. C. Davis that the 'Ranters' were inventions of seventeenth century conservatives and twentieth century Marxists. | Radicalism in the eng Revolution |
| 0019e | What is the nature of Gerrard Winstanley's religious outlook in The New Law of Righteousness 1649? | " |
| 0019g | Why were Chinese luxury goods of so much fascination to seventeenth and eighteenth century Europeans? | Luxury and Pleasure |
| 0019h | What were the implications of the eighteenth and late nineteenth century luxury debates? | " |
| 0029e | Do Western understandings of Asia continue to be informed by Orientalist assumptions? | Historiography |
| 0029f | Does Ranke deserve the title 'father of scientific history'? | " |
| 0029h | What factors account for the failure of the Church Missionary Society's Wellington Valley Mission? | Antipodean Encounters |
| 0029o | In what ways and to what extent did an Enlightenment world-view shape life in colonial New South Wales from c. 1788-1850? | " |
| 0040c | Are you convinced by the view of the Levellers on franchise reform as expressed by McPherson in his Political Origins of Possessive Individualism? | Radicalism & the eng Revolution |
| 0040e | In what ways did Braudel conserve and in what ways contradict the Annales tradition? | Historiography |
| 0042c | Does Ranke deserve the title 'father of scientific history'? | " |
| 0042d | Do Western understandings of Asia continue to be informed by Orientalist assumptions? | " |
| 0144d | 'Braudel betrayed the Annales tradition.' Discuss. | " |
| 0144e | What part did the factors outlined in Weber's Protestant Ethic play in explaining the divergent economic development of Western and Eastern civilisations | Historiography |
| 0244e | Were Proletarian Factory Workers less likely than Artisans to Participate in Industrial and Political Militancy - and if so, why? | Workers and Society in Europe 1870-1945 |
| 0244f | Was the Italian Labour Unrest of 1919-20 defeated more because of divisions on the left than because of the fascist counter-offensive? | " |
| 0244g | What Reasons were there for Growing Radicalism in the Russian Labour Movement in the Years Before 1914? | " |
| 0252c | 'Madame de Pompadour was popular in Versailles, unpopular in Paris.' Discuss. | Enlightenment and Revolutionary Paris 1750-99 |
| 0252r | With particular reference to the work of Jacques-Louis David, how were the principles of the French Revolution anticipated, reflected and promoted in French painting, 1784-1794? | " |
| 0255g | TO WHAT EXTENT WERE LENIN'S POLICIES GUIDED BY A CONSCIOUS STRATEGY FOR THE TRANSITION TO SOCIALISM BETWEEN 1917 & 1921? | Russian Revolution |
| 0255h | Does Ranke deserve the title 'father of scientific history'? | Historiography |
| 0314i | To what extent did relations with the Soviet Union and American Left shape the character of the Mexican Revolution? | Race, History and Nation in Mexico |
| 0318e | Constantin von Neurath (1873-1956) | Eastern Europe in Crisis |
| 0318f | The rise and fall of the Third Reich | " |
| 0318g | Leopold von Ranke (1795-1886) | Historiography |
| 0380b | Analyse the relationship between prices and theft prosecutions in Elizabethan and Jacobean England. | English Communities in Crisis |
| 0380c | Roy Porter describes private asylums as both 'running sores of scandal' and 'sites of therapeutic innovation.' Do you agree with this dual imagery? | Madness and Society: from Bethlem to the Present |
| 0380d | Could an argument be made that mental illness is socially constructed with respect to hysteria or neurasthenia? | " |
| 0391c | What impact did the hunger strikes in 1981 have on the politics of Northern Ireland? | The Troubles 1968-1998 |
| 0391e | What have been the major historiographical arguments about China's 'tribute system'? | China Encounters |

POL/IR-X

| Essay | Essay title | Topic |
|------------|---|--|
| PIR3_002_C | Account for the tensions between 'Westbindung' and 'Ostpolitik' in West German foreign policy | The politics of Modern Germany |
| PIR3_002_D | What factors and conditions worked to undermine efforts to forge a unified region-wide Arab state in the region before 1948? | Middle Eastern Politics |
| PIR3_002_E | The EU needs a European Army. Do you agree | Defence in the post-Cold War Era |
| PIR3_003_P | To what extent can the current Sino-US relationship be described by the policy of containment? | " |
| PIR3_003_Q | Is Hegel right to think that freedom is actualized in the ethical life of a modern state? | Issues in Democratic Theory |
| PIR3_003_R | Does international law only reflect the will of powerful states? | Advanced Readings in Global Studies |
| PIR3_003_S | Describe and critically appraise the concept of web 2.0. | The Politics of the Internet and the Information Society |
| PIR3_003_T | Why for Freud can there be no genuinely free society of either a Hegelian or Marxist (or other) type? | Issues in Democratic Theory |
| PIR3_003_U | Should we, in the words of Luttwak, forget about peacekeeping and "give war a chance"? | Defence in the post-Cold War Era |
| PIR3_004_Q | Why have women in South Asia been complicit in social and political movements that ultimately constrain them? | The Making of Modern South East Asia |
| PIR3_004_R | To what extent can Bayart's "Politics of the Belly" be used to understand the State in Africa? | The Politics of Africa |
| PIR3_004_S | Did 9/11 have any impact on our understanding of self-defence? | Defence in the post-Cold War Era |
| PIR3_004_T | What role has the army played in Pakistani politics? | The Making of Modern South East Asia |
| PIR3_004_U | Should we, in the words of Luttwak, forget about peacekeeping and "give war a chance"? | Defence in the post-Cold War Era |
| PIR3_005_N | What reasons do you think accounted for the SPD's emergence as a party of government on the federal level in the 1960s? | The Politics of Modern Germany |
| PIR3_005_O | Nationalism in the Middle East began to assume a definite form after the Young Turks came to power. Why? What were the nationalist movements that arose during this time? Who led them, what were their aims? | Middle Eastern Politics |
| PIR3_005_P | Social discontent and nationalism were strong forces behind the Egyptian revolution and the rise of Nasser. How did these two forces combine to spark upheaval and revolution? To what extent was Nasser's socialism and pan-Arabism an adequate response towards addressing the popular concern? | " |
| PIR3_005_Q | International courts have never worked and they never will. Do you agree? | International Law |
| PIR3_005_R | The case of the 2003 Iraq war shows why pre-emptive self-defence should be illegal in international law. Do you agree? | " |
| PIR3_006_M | What is the relationship between resistance and identity? | Resistance to Global Governance |
| PIR3_006_N | To what degree and in what ways has colonialism shaped African politics since the 1960s? | The Politics of Africa |
| PIR3_006_O | 'Structural Adjustment Programmes were designed to keep the continent economically subservient to the west.' Why might some people make this argument? How far do you agree with it? | " |
| PIR3_006_P | Why is it so difficult to successfully prosecute those responsible for the crime of genocide? | International Law |
| PIR3_006_Q | Are International Courts Effective? | " |
| PIR3_006_R | The Role of International Organisations in World Politics | Advanced Readings in Global Studies |
| PIR3_007_I | How does the discourse of the Occupy Wall Street protests relate to Foucault's conception of resistance? | Resistance to Global Governance |
| PIR3_007_J | Can the ideas of governmentality be used to understand contemporary challenges to international security? | " |
| PIR3_008_Q | What does the tactic of 'Kettling' tell us about the interplay between Resistance and Power? | " |
| PIR3_008_R | 'Is there a place for resistance in a Foucauldian account of politics?' | " |
| PIR3_008_S | 'What does the "world polity" thesis tell us about the relationship between nation-states under conditions of globalization?' | Advanced Readings in Global Studies |
| PIR3_008_T | 'What is the dialectic of lordship and bondage (aka master and slave) and what does it say about self-consciousness and freedom?' | Issues in Democratic Theory |
| PIR3_008_U | What were the forces that gave rise to the Internet? | The Politics of the Internet and the Information Society |
| PIR3_008_V | What is Nietzschean genealogy? Is it an effective form of critique? | Issues in Democratic Theory |

POL/IR - BAWE

| Essay | Essay title | Topic |
|-------|--|--|
| 0137l | On Capital Punishment | Issues in Political Theory |
| 0137m | Threat of Islam or threat to Islam? Critically assess the conflict between Islam and the West. | Middle East Regional Relations |
| 0137n | Transcending Victimhood: A Critical Analysis of Women in Ethnic Conflict | Ethnic Conflict and Political Violence |
| 0234j | What, if anything, has membership of a political community in common with membership of a family? Can this tell us anything about our obligations? | Issues in Political Theory |
| 0244j | To What Extent are the Grand Theories of Integration Useful Tools for Analysing European Union Policymaking? | The Politics of European Union Policy Making |
| 0244k | Should Surfers be Fed? A Critical Examination of Arguments for and Against Unconditional Basic Income. | Issues in Political Theory |
| 0244l | Are there 'Two Concepts of Liberty'? Distinguish Some Different Contrasts which have been made between Positive and Negative Liberty. | " |
| 0244m | To what Extent has the EU Contributed to Global Environmental Protection? | The Politics of European Union Policy Making |
| 0324b | Does Rawls's conception of "justice as fairness" provide a superior alternative to utilitarian conceptions of justice? | Issues of Political Theory |
| 0399e | To What Extent is 'New Institutionalism' useful when analysing E.U Processes? (Do Institutions really Matter?) | European Union Policy Making |
| 6180b | How far has diplomacy changed since the end of the Cold war? | Foreign Policy Analysis |
| 6180d | Why has there been a rise in regionalism in the international economy? | Politics of the International Economy |
| 6180k | Examine diplomatic methods used by Russian Fed in the conduct of its FP. In what ways, if at all, does it reflect dom interests? | Foreign Policy Analysis |

Appendix 2: Essay titles categorised according to focus

| Focus on | HIST-X | HIST-BAWE | POL/IR-RH | POL/IR-BAWE |
|--|---|--|--|---|
| the specific (actors, contexts or phenomena) | 001_X; 001_Y; 001_Z; 009_A; 009_B; 009_C; 009_D; 009_E; 009_F; 010_A; 010_B; 010_C; 010_D; 010_E; 010_F; 011_R; 011_T; 011_U; 011_V; 011_W; 012_A; 012_B; 012_C; 012_D; 012_E; 013_G; 013_H; 013_I; 013_J; 013_K | 0019; 0019d; 0019g; 0019h; 0029f; 0029h; 0029o; 0040e; 0042c; 0042d; 0144d; 0144e; 0244e; 0244f; 0244g; 0252c; 0252r; 0255g; 0255h; 0314i; 0318e; 0318f; 0318g; 0380b; 0380c; 0380d; 0391c; 0391e | 002_C; 002_D; 002_E; 004_Q; 004_S; 004_T; 005_N; 005_O; 005_P; 005_R; 006_N; 008_U; 003_P | 0244j; 0244m; 6180b; 6180d; 6180k |
| general concepts applicable across a range of instances | 001_P; 001_Q | 0019e; 0029e; 0040c; 0042d; 0380d; | 003_Q; 003_R; 003_S; 003_T; 003_U; 004_R; 004_U; 005_Q; 006_M; 006_O; 006_P; 006_Q; 006_R; 007_I; 007_J; 008_Q; 008_R; 008_S; 008_T; 008_V; | 0137l; 0137m; 0137n 0234j; 0244k; 0244l; 0324b; 0399e |
| explanation of cause-effect processes | 001_X; 001_Y; 001_Z; 009_D; 009_E; 009_F; 010_B; 010_C; 010_E; 010_F; 011_R; 011_T; 011_U; 011_V; 011_W; 012_C; 013_G; 013_H; 013_I; 013_J; 013_K | 0019g; 0019h; 0029e; 0029h; 0029f; 0029o; 0042c; 0144e; 0244e; 0244f; 0244g; 0255g; 0255h 0314i; 0318f; 0318g; 0380b; 0391c; | 002_C; 002_D; 004_Q; 004_S; 004_T; 005_N; 005_O; 005_P; 006_N; 008_U; | 0244m; 6180b; 6180d; 6180k |
| ideas, theories or issues | 001_P; 001_Q; 012_B; 012_D; 012_E | 0019d; 0019e; 0029e; 0040c; 0042d; 0380d | 002_E; 003_Q; 003_S; 003_T; 003_U; 004_R; 004_S; 004_U; 005_Q; 005_R; 006_M; 006_O; 007_I; 007_J; 008_Q; 008_R; 008_S; 008_T; 008_V; | 0137l; 0234j; 0244j; 0244k; 0244l; 0324b; 0399e |
| evaluation of the nature of phenomena. | 009_A; 009_B; 009_C; 010_A; 011_T; 011_U; | 0019d; 0029f; 0040c; 0042c; 0144d; 0252c; 0252r; 0255h; 0318e; 0318g; 0380c; 0391e | 003_P; 003_R; 006_P; 006_Q; 006_R; | 0137m; 0137n 0234j; 0399e |

Appendix 3: PIR3_008_R – phraseologies highlighted

‘Is there a place for resistance in a Foucauldian account of politics?’

Foucauldian accounts resistance stem from a re-conceptualization of power in opposition to ideas that power is located solely in the state apparatus, is the method of subjugating a population is the possession of a government or generally acts from rulers to the ruled in a downward action. Instead it is seen as being something existing everywhere, originating from innumerable points and moving in a network between numerous subjects. Resistance is seen as being not exterior to but within the power network itself, resistance is not against power but inherent in relationships of power. Using Foucault’s re-conceptualization of power and resistance I shall argue against Marxist and Anarchist ideologies of resistance that characterise power as being the possession of a state and political resistance as being either the seizure or abolition of power centred within a state. I shall make the case that forms of resistance should learn from Foucauldian notions of power and resistance, recognise the true nature of power as inescapable and embrace forms of resistance that are heterogeneous in origins and local in target, directed towards interstices of dominating power relations.

Firstly let us examine the traditional theories of Marxism and Anarchism, how they conceive of the state, revolutionary activity and how they see the role of resistance, beginning with Marxism. For Marxism power is located solely within the state and in the hands of the bourgeoisie. As Foucault puts it, Marxism holds a juridico-economic conception of power as ‘a right which can be possessed in the way one possesses a commodity’ (Foucault. 2003: 13). The bourgeoisie, the ‘class of modern

Capitalists, owners of the means of social production and employers of wage labour' (Marx. 1967: 79) are the sole possessors of power.

The workers and the capitalists sit in a relation of constant exploitation and domination. Power is exercised purely in a downward fashion, always by the capitalists upon the workers, the two groups stand 'in constant opposition to one another, carried on an uninterrupted, now hidden, now open fight' (Ibid). Relations of dominations exist historically and presently between a majority of the population and a minority of rulers. These relations are one of constant antagonism, with the dominate class seeking to exploit the lower class. The classes are not limited to capitalist and worker, but to other historical oppressors and oppressed, such as feudal lord and serf.

Furthermore, this struggle that exists between two classes in society is, in and of itself, the sole struggle existing. All struggles, or shall we say acts of resistance, can be homogenised as being part of the sole struggle against (or in defence of) capitalism – as Marx eminently put it: 'The history of all hitherto existing society is the history of class struggles' (Ibid). This homogeneity is reflected in the notion of Communist revolution, which desires the single, unitary overthrowing of capitalism:

'The Communists disdain to conceal their views and aims. They openly declare that their ends can be attained only by the forcible overthrow of all existing social conditions. Let the ruling class tremble at a Communistic revolution. The proletarians have nothing to lose but their chains. They have a world to win. Working men of all countries, unite!' (Ibid: 120-121).

The struggle therefore is for the seizure or ‘conquest of political power’ by the proletariat’ (Ibid: 95). A Marxist conception of resistance is unitary in its goal, which is the ‘dictatorship of the proletariat’, the seizure of ‘power’ (state, economy) from those who own it, by those who possess no power.

Finally, Marxism is a dialectical philosophy. The progression of history is a progression of modes of production, a progression which Marx traces through periods such as slave based societies, to feudalism, to capitalism, and of course eventually to communism. Fundamentally, when concerning dialectics, the progression is one of an internal movement, driven internal antagonism. These antagonism in Marx’s dialectical method is the interplay between the productive forces and the relations of production, with synthesis of and progression occurring when the two become incompatible, when the relations of production serve to hamper the productive forces and prove their ‘fetters’. Just as dialectical history resolved the antagonisms of feudalism by progressing into capitalism, capitalism is presently undergoing ‘A similar movement [...] before our own eyes’ (Ibid: 85). In this movement are the antagonisms between proletarian and bourgeois played out, the struggle between classes, in other words the Marxist conception of political resistance. Resistance therefore is not an individual act but a pre-determinate occurrence, a product of a dialectical movement.

Now to Anarchism, and its own approach to the questions of power and resistance.

Fundamentally speaking, Anarchism is:

‘A fundamental critique and rejection of political authority in all its forms. The critique of political authority – the conviction that power is oppressive, exploitative and dehumanizing – may be said

to be the crucial politico-ethical standpoint of anarchism'

(Newman. 2005: 33).

For Bakunin, for example, the state is viewed as 'simply an instrument of political force'. Physical coercion had always been the backbone of the state' (Saltman. 1983: 54). The states control of power is centralized and exclusive, it's 'coercion had been subsumed into a set of "well-regulated and systematized" institutions that centralized the state's control over all' (Ibid). Anarchists therefore locate 'the fundamental oppression and power in society in the very structure and operation of the state' (Newman: 34).

As power is located solely in the state, and the state is the oppression of man, then the goal for anarchists is to abolish the state, therefore abolishing power. As Bakunin puts it: 'Abolition, dissolution and moral, political, legal, bureaucratic and social bankruptcy of the custodial, transcendental, centralist State' (Bakunin. 1973: 66).

'The strict Manichean separation between humanity and power on which classical anarchism bases itself would be seen, from a Nietzschean perspective, as an expression of the Apollonian illusion [...] of a life-world without power. (Newman: 41).

Also crucial to Anarchism, as with Marxism (though less discussed) is a dialectical logic, as Newman points out, Anarchism 'subscribes to a dialectical logic, according to which the human species emerges from an 'animal-like' state and begins to develop moral and rational faculties' (Newman: 38). The dialectics in anarchism begin with a human essence; the makeup of which comprises a natural pre-disposition towards cooperation and mutual aid, rather than competitiveness.

This human essence, however, is denied when it, natural order, is put under **the authority of states**, artificial order.

‘Anarchism **may be understood**, then, **as a struggle between these two forms of authority** [artificial order/natural order]. **This struggle** is part of a dialectical process in which the subject develops to **a state of full humanity**’ (Newman: 37).

Now we have considered both Marxism **and** Anarchism, we have exposed that; for Marxism **and** Anarchism power exists solely in the hands of the state **and** functions downwards; both subscribe to a dialectical logic, of which, the Anarchist dialectics rest on **a conception of human essence**, **and** Marx’s dialectics rest on **the drive of antagonistic forces**; resistance is a resistance to either negate or seize **the power of the state**; resistance should be, or is, homogenous, having **its aim as the unitary and climatic revolution against capitalism**.

Let us now consider **the Foucauldian conception of power and resistance** **and** how it stands contra **the position of Marxism and Anarchism** outlined above. For Foucault power is conceived of not ‘as **a group of institutions and mechanism that ensure the subservience of the citizens of a given state**’ (Foucault. 1976: 92), by this Foucault means to say **that power is not strictly the state and it’s mechanism, but something more complex**. It is also not ‘**a general system of domination** exerted by one group over another’ (Ibid), instead it is seen as:

‘**the multiplicity of force relations** immanent in the spheres in which they operate **and** constitute their own organization; **as the process** which, through ceaseless struggles **and** confrontations, transforms, strengthens or reserves them; **as the support which**

these force relations find in one another, thus forming a chain or system' (Ibid).

Power, therefore is manifold in that it exists as a numerous relations of force between a multiplicity of subjects. These relations of force, power relations, are fluid, they struggle and change over time as force is directed at different localities or targets. Some relations serve to strengthen others, a commonality serving to make links between power relations, through which systems or congealment occurs, creating dominating power structures. Furthermore force relations are:

‘The strategies in which they take effect, whose general design or institutional crystallization is embodied in the state apparatus, in the formulation of the law, in the various social hegemonies’ (Ibid. 92-93).

This ‘crystallization’ is the formation of states, comprising of laws and apparatuses – they are part of the definition of power but not the sole proprietor of it.

Examined in more detail we can see how Foucauldian notions of power expose the simplicity of previous conceptions. For both Marxism and Anarchism power is a commodity that can be possessed by a monarch or president, can be traded at the passing of a law or the winning of an election, can be usurped in a revolution, or can be simply done away with. Yet Foucault asserts the opposite; that ‘Power is not something that is acquired, seized or shared’, it is ‘exercised from innumerable points, in the interplay of [...] mobile relations’ (Ibid. 94). Power is not static within a government but mobile through relations. It’s exercised innumerable in the sense that all people in interpersonal relations, all groups in society, all states

in the international sphere, exercise it when they interact with those in the network of relations surrounding them.

Whilst Marxism and Anarchism both hold **that** power works downwards, power can be seen to come 'from below'; that is, there is no binary and all-encompassing opposition between rulers and ruled at the root of power relations' (Ibid). Worker and capitalist do not exist solely in opposition to one another; the actors involved are considerably more numerable. 'The manifold relationships of force that take shape and come into play in the machinery of production, in families, limited groups, and institutions, are the basis for wide-ranging effects of cleavages that run through the social body as a whole' (Ibid).

'Where there is power, there is resistance' (Ibid: 95). Resistance cannot be understood simply as being external or opposed to power, it is itself a part of the force relationship previously discussed. 'These points of resistance are present everywhere in the power network', it is wrong to assume all resistance can be homogenized to form a single revolutionary movement. Instead moments of resistance are acts directed at local regimes of power. Newman, citing Laclau and Mouffe, points out how 'the struggles of workers and artisans in the nineteenth century tended to be struggles against relations of subordination generally, **and** against the destruction of their organic, communal way of life'. These struggles did not conform to the homogenising Marxian notion **of** workers 'embracing the forces of capitalism in order to radicalize it' (Newman: 48). Worthy quoting at length is Foucault's criticism of the notion of homogenous resistance and his assertion of heterogeneity:

‘Hence there is no **single locus of great Refusal**, no **soul of revolt**, **source of all rebellions**, or **pure law of the revolutionary**. Instead there is a **plurality of resistances**, **each of them** a special case: **resistances that are possible, necessary, improbable; others that are spontaneous, savage, solitary, concerted, rampant, or violent; still others that are quick to compromise, interested, or sacrificial**; by definition, they can only exist in **the strategic field of power relations**’ (Foucault. 1976: 95-96)

But what area specifically and practically does Foucault give to **resistance as an act against domination**? For Foucault, **the primary act of resistance** is what he calls ‘genealogy’, which is:

‘[An] attempt to desubjugate historical knowledge, to set them free, or in other words to enable them to oppose and struggle against the coercion of a unitary, formal, and scientific theoretical discourse’ (Foucault. 2003: 10).

When referring to ‘subjugated knowledges’ Foucault refers to:

‘**A whole series of know ledges that have been disqualified as nonconceptual knowledges, as insufficiently elaborated knowledges, knowledges that are below the level of erudition or scientificity**’ (Ibid: 7).

Foucault is attacking, what he refers to as, **the domination of scientific discourses**. He sees this is a tendency to legitimize or delegitimizes speakers according to scientificity. So for example, Marxists might want to assert **the scientific nature of**

Marxism, Foucault sees this attempt not as an attempt to assert Marxism's 'rational structure and [...] its propositions are therefore the products of verification procedures', but instead as an act that has the result of disqualifying other types of knowledge, and legitimizing its own discourse. By gaining the label 'scientific' Marxism seeks not rational recognition but legitimacy as a speaker of a scientific discourse, of a scientific truth. The claim to be scientific is an attempt to benefit from 'the power-effects that the West has, ever since the Middle Ages, ascribed to a science and reserved for those who speak a scientific discourse' (Ibid: 10), its effect is to subjugate other knowledges that cannot attain the exclusionary title of 'scientific discourse'.

So for example Foucault established the 'Groupe d'Information de Prisons', an attempt to give a voice the prison inmates who had been excluded from debate because they failed to meet the criteria of legitimate speaker or scientific discourse. This group served as a form of resistance against the domination of scientific discourse:

'If the discourse of inmates or prison doctors constitutes a form of struggle, it is because they confiscate, at least temporarily, the power to speak on prison conditions – at present, the exclusive property of prison administrators and their cronies in reform groups' (Foucault and Deleuze. 1972)

Newman, by applying Foucauldian models to Anarchism gives us a good example of what forms of resistance have to gain from a post-structuralist reading of resistance. Firstly he criticizes classical anarchism's dialectical basis and secondly its conception of a human essence. Consider again Foucault's understanding of power

as something ‘exercised from innumerable points, in the interplay of non-egalitarian and mobile relations’ (Foucault. 1981: 94). Furthermore, as Newman puts it, identity has a ‘discursively constituted nature’ (Newman. 2005: 46). Our identities are the effects of power relations, post-structuralism stresses the ‘structural instability and un-decidability of our identity’ (*Ibid*: 46). The conception of a human essence therefore is shattered when we consider the way our identity is formed and its fluid and changeable nature. The dialectical method too is flawed. The notion of a determinate path of dialectical history is entirely implausible. History is not the unfolding of a rational truth, ‘Rather, history is a haphazard and unpredictable series of ruptures, discontinuities and events’ (Newman. 2005: 47). *The way in which* power relations stress and pull at various locations in various directions means history is not pre-determined, but entirely random.

Newman’s conclusion is **that** contemporary anarchism must ‘distance itself from its classical foundations in the dialectic and positivist and essentialist ideas of Enlightenment humanism, and assert instead [...] political contingency and heterogeneity’ (*Ibid*: 49). Radical politics must embrace ‘a multiplicity of different identities and struggles’, particularly those on a local level at the ‘interstices of power’ (*Ibid*). Newman coins the term post-anarchism to describe **this new Foucauldian movement**. **This movement** would be ‘a series of politico-ethical strategies **that** are aimed at the deconstruction of authority, exposing domination and coercion behind institutions’ norms and practices’ (*Ibid*). Radical politics generally, **and** not just anarchism, would benefit from an adoption of a heterogeneous approach to resistance. Heterogeneity, in its local and specific origins, diversifies a movement and makes it more autonomous.

Therefore *it can be seen that* power and resistance undergo a critical re-conceptualization in the works of Foucault. For Foucault specifically, resistance is, in one instance the desubjugation of historical knowledges. In the new light of this re-conceptualization, Marxism and Anarchism, in their classical varieties suffer from a tendency to simplify the nature of power and homogenize resistance. In the same vein as the Post-Anarchism of Newman, radical politics generally must embrace the Foucauldian account of power and resistance and seek heterogeneous, local critiques of power.

Appendix 4: Analysis - of – pooled sample

| Concordance | Set |
|---|-----|
| disfigured without the appendages OF Bourbon monarchy, thus indicating that | 1 |
| provide the contextual background OF each ethnic conflict to the extent the | 1 |
| means to his sharing in the bounty OF China, and nothing more. Primarily the | 1 |
| the French Revolution the culture OF the artisan was nothing if not social. | 1 |
| increase in the productive forces OF the country...all other considerations | 1 |
| priority or 'consumption OF the fruits OF the State'. In a radio and television | 1 |
| in Weber's analysis for the future OF economic progress. Seeing the spirit C | 1 |
| the social and cultural landscape OF the ancien r gime, and replaced it wi | 1 |
| OF abortion has entered the lexicon OF feminist struggle through a very diffe | 1 |
| ere in the political organisations OF the continent. It appears that the 'gc | 1 |
| desperate minority for the riches OF the State". Therefore corruption is r | 1 |
| p. 339. Extending from the rights OF the individual to exact punishment fro | 1 |
| e, in place OF the multiple skills OF a single man, the pooling OF the techni | 1 |
| ces, the uneasy theme and structure OF the Brutus is probably more likely to | 1 |
| . While the geopolitical structures OF the world remain fundamentally Westpha | 1 |
| ne overriding geopolitical tensions OF the early Cold War, the economic reali | 1 |
| g this question because the themes OF this essay, such as being labelled by | 1 |
| ns and world poverty are new topics OF the international agenda. The Kyoto Pr | 1 |
| rding to the requirements and logic OF the hour.' La M diterran e is not, t | 1 |
| ere, the SPD under the personality OF Willy Brandt came to run the country f | 1 |
| role OF women and the personality OF the Statesmen. Bayart's thesis illustr | 1 |
| or lack thereof. Thereby, the role OF institutions becomes a more ideologica | 1 |
| rance, a say in the future workings OF the Council. The same utility can be a | 1 |
| an ICC and the greatest achievement OF the international criminal court has b | 1 |
| revolutionary model as a manifesto OF the revolution in Russia and its Marxi | 1 |
| his treatment for gout was the foot OF a green toad mixed with other animal b | 1 |
| are like patterns on the surface OF water, which appear fixed when seen fr | 1 |
| levels. For example, for occupation OF the criminal vagrant is rarely stated: | 1 |
| it not to ignore the artistic style OF Ranke's history - 'whereas the princip | 1 |
| into an internationalised theatre OF the Cold War conflict. According to Re | 1 |
| urate concept that forms the basis OF a conflict with the "West". More impor | 1 |
| le in society even if the position OF the least advantaged is not maximined | 1 |
| in genuinely observant OF the rites OF religion') his zealous promotion OF Re | 1 |
| o equip ourselves with the weapons OF democracy from the arsenal OF democrac | 1 |
| at, ultimately, land is the source OF wealth. This is true for both feudal a | 1 |
| istance embodied in the discourses OF the New York protests alone. It uses a | 1 |
| edges." In this way, the discourse OF OWS most certainly relates to Foucault | 1 |
| alist ideology . The leftist milieu OF the early Nasser regime was also commu | 1 |
| asure "the qualitative composition" OF the proletarian component. To protect | 1 |
| t system that linked the outskirts OF the great city to her centre, providir | 1 |
| were seen as being at the very top OF the chain. However, there was a great | 1 |
| to Orientalism as mere components OF the homogenising discourse. He has cla | 1 |
| tions - the imperialist aspiration OF Germany was no different one must add. | 1 |
| deeply engrained in the attitudes OF the elite that it was a natural export | 1 |
| colonists and the dreamtime beliefs OF the Aborigines, had disastrous consequ | 1 |
| h the basics concerning the nature OF female medical aid and what it involve | 1 |
| ably had more to do with the nature OF the challenges he faced than his own p | 1 |
| however, what determined the nature OF relations between the two was the char | 1 |
| international law, the entire nature OF the post-war international settlement | 1 |
| lation, thus distorting the nature OF the metropolis as vast stretches OF va | 1 |

| | | |
|----------------------------------|---|---|
| However, the most important part | OF his legacy as the last pagan emperor | 1 |
| s, that tended to favour members | OF the great families. It would have also | 1 |
| ers... My liberty is the liberty | OF everybody; for i am really free, free | 1 |
| t can use the predictable crisis | OF the current orange-type regimes to reg | 1 |
| ts, although rich in the details | OF events, lack. The crusade was launche | 1 |
| be a dispute between the authors | OF "Winstanley: A Case for the Man as He | 2 |
| heimer was a great driving force | OF the Manhattan project and his moral sc | 2 |
| egration and an active supporter | OF German involvement and enmeshment in t | 2 |
| is by no means a principal cause | OF poverty, rather a more plausible expla | 2 |
| volution, and David's own choice | OF style and subject within his revolutio | 2 |
| e OF ecclesiastical condemnation | OF their preaching activities. The Walder | 2 |
| . 1983: 54)>. The states control | OF power is centralized and exclusive, it | 2 |
| OF policies leading to creation | OF institutions at a regional level. It s | 2 |
| ed' form, suggesting deportation | OF undesirables to the East. (Speech made | 2 |
| distinctions through the display | OF consumer objects." Here Ritzer is refe | 2 |
| ty's proposals for the exclusion | OF servants and alms-takers, whether they | 2 |
| is "written in" to the exercise | OF power. If this is true, then the agenc | 2 |
| were made to ban the exhibition | OF the painting in the 1789 Salon on the | 2 |
| tations, such as the expectation | OF being available day and night, OF what | 2 |
| lries. OF course, identification | OF a common enemy can help to forge a sha | 2 |
| to evidence OF the incorporation | OF those principles soon to become Revolu | 2 |
| ady been made with the launching | OF the ambitious and controversial Earth | 2 |
| g a member far outweigh any loss | OF sovereignty that will take place from | 2 |
| sire. Typical examples OF a loss | OF negative freedom can be found in situa | 2 |
| ugh his re-awakening and mastery | OF nature the slave is eventually seen by | 2 |
| te in 1900 that the presentation | OF 'facts, facts, facts... The truth, all | 2 |
| the trend towards the promotion | OF Turkish racial superiority intensified | 2 |
| ntinued influence and protection | OF their economic interests in Mexico. De | 2 |
| g tool as well as for protection | OF the Czech population, an undetectable | 2 |
| ism, Weber refers to the pursuit | OF profit by means OF continuous, rationa | 2 |
| ogical and cultural reproduction | OF the nation". With that, women assert a | 2 |
| has hence argued for the reserve | OF the term 'tribute system' for the late | 2 |
| sade until after the subjugation | OF Carcassonne, by which time revelations | 2 |
| hedral to aid in the translation | OF the Bible. In essence this desire to s | 2 |
| 6 which led to "the unionization | OF some 9,500 'militant' metalworkers in | 2 |
| tains the prohibition on the use | OF force in Article 2. However there are | 2 |
| slave morality. Post-modern uses | OF genealogy explore its ability to expos | 2 |
| ays he was called 'The adornment | OF Islam, the king helped by God, the hel | 2 |
| t figure in the Czech experience | OF occupation during his time in office a | 2 |
| as part OF the lived experience | OF the early Revolution, the people had b | 2 |
| ve liberty, and not some mixture | OF the two. He is also very careful to di | 2 |
| in that it would reduce the fear | OF extreme deprivation, or OF uncertainty | 2 |
| ension, 'getting rid OF the fear | OF damnation." Secondly, MacKinnon argues | 2 |
| urSint Forum . This amalgamation | OF overlapping intelligence gathering mec | 2 |
| gimes based on the "sovereignty" | OF the people" rather than "God's right t | 3 |
| ashion, which "became a function | OF knowledge rather than wealth, and gave | 3 |
| lls this the "semiotic" function | OF goods. Michael Spindler, a historian C | 3 |
| paper is to investigate the role | OF an international institution as a stag | 3 |
| ed doubts about the central role | OF environmental policy in the future pro | 3 |
| there was a complete break-down | OF effective government. In 1055 the Sel | 3 |

| | | |
|----------------------------------|--|---|
| power for himself on the death | OF Nur al-Din and his son Malik-Shah. The | 3 |
| vernment, resulted in the death | OF between 700 and 2,500 and provoked Pre | 3 |
| ss OF confusion with the death | OF their leader Simon de Montfort. This r | 3 |
| se the import OF what the death | OF god means has not reached them yet, as | 3 |
| erequisite for the development | OF socialism. At the time when this was v | 3 |
| t that there was a slowing down | OF time. The Annales increasingly favoure | 3 |
| Ericson observed the evolution | OF this trend since the 21st century, whi | 3 |
| ons contributing to the failure | OF pan-Arabic statism is imperialism as e | 3 |
| ensianism, as after the failure | OF the Gregorian reforms in the previous | 3 |
| ory Workers, and the Formation | OF the French Working Class, p. 53. Roger | 3 |
| n world was the intensification | OF the Cold War, specifically the Korean | 3 |
| non-state actors and intrusion | OF foreign states in previously domestic | 3 |
| horrible. As such, by the onset | OF the middle ages the image OF magic was | 3 |
| ates is for mediation, rallying | OF nations, and generating awareness to p | 3 |
| ethno-national groups. The rise | OF nationalist ideology in ethno-national | 3 |
| thereby contributed to the rise | OF regional projects. The conflict ceased | 3 |
| factors gave rise to the surge | OF recent regional projects and motivated | 3 |
| erable points, in the interplay | OF non-egalitarian and mobile relations' | 3 |
| was the charismatic leadership | OF Willy Brandt. The fact that, even duri | 3 |
| 22-23, page 21: "the existence | OF a positional duty...is a morally neutr | 3 |
| ertheless illustrates the fears | OF the church about unlicensed preachers | 3 |
| to directly coincide with those | OF the nations which had previously explo | 3 |
| ry was a science. The scenario | OF 'all things being equal' does not exist | 4 |
| Muslim world through the "prism | OF violence and terrorism" which has led | 4 |
| ood most, promulgating an ideal | OF histoire totale. Jean Gl  nissou, 'Fra | 4 |
| ethods. Taylorism and the ideal | OF ergonomics seemed to provide the perfe | 4 |
| ictures, the more mobile forces | OF social and economic cycles;   v  n  me | 4 |
| litically transformative policy | OF Ostpolitik . Another significant alter | 4 |
| resources it reduces the problem | OF having to work out and favour a partic | 4 |
|    satisfaction against the risk | OF pain or social punishment. Therefore t | 4 |
| shaped by historical traditions | OF tribal ruling and patronage. This clar | 4 |
| essarily dissimilar to the case | OF the political community. The differenc | 4 |
| on. New journalism, or the case | OF Wiki Leaks , are examples which suppor | 4 |
| which earned him the reputation | OF 'father OF scientific history' in the | 4 |
| all animated by the common aim | OF enhancing individuals' occupational fr | 4 |
| nt OF divergence from the peace | OF normality. 'We shifted the focus OF ou | 4 |
| tion" and "endless relativism" | OF postmodernism at the same time. Prakas | 4 |
| es OF knowledge. From each body | OF knowledge, information is extracted ar | 4 |
| t was initially agreed that 70% | OF the cost would be covered by Khrushche | 5 |
| p Braudel often did not use all | OF the Annaliste tools, yet no one to dat | 5 |
| ith EU law." As a result OF all | OF these difficulties there are still hug | 5 |
| ne history and function OF each | OF these elements in this essay, especial | 5 |
| t devoured more than one fourth | OF Pakistan's annual budget leaving very | 5 |
| l-Mans  r. However, the majority | OF complaints from his practical examinat | 5 |
| the protection OF the majority | OF 'useful' Czechs, but exposed the rest | 5 |
| Industrial Revolutions, and many | OF their answers coincided with what Rank | 5 |
| ' and are given to some members | OF society over others. Again there is a | 5 |
| Britain were able to block most | OF France's ideas with regard to Germany, | 5 |
| Felt across Europe and the rest | OF the world. The EU's website even gives | 5 |
| levels OF GDP growth with some | OF them being negative. During the disast | 5 |

s not likely that the general body OF wage-earners would be intimidated, but 5
er. These realms cover all aspects OF a citizen's life and allow him to be v 5
ower, states gain a greater degree OF relative power and in this sense exter 5
ecurity Strategy, that is the lack OF articulated military capabilities . Wi 5
therefore, our duty to avoid lack OF solicitude.' One OF Islam's greatest c 5
consequence, states develop a lack OF will to support prosecutions for genoc 5
ntemporary society. The advantages OF progress often translated into hopes f 6
Ireland. The contradictory nature OF Britain's foreign policy was also expc 6
e strictly security-centric nature OF this context, 'techniques' OF governme 6
1213 epitomises the complexities OF the Albigensian crusade: on the side OF 6
also commented on the limitations OF secular work as "secular presuppositio 6
The Levellers held a diverse range OF opinions and at this point in the move 6
"the winds OF change". The variety OF jubilant gestures performed by the del 6
First World War. The effectiveness OF the Soviet as a body for socialist rep 6
nto conspicuous waste and excesses OF 'old luxury', and taste becomes once a 6
bomb.' The excitement and intrigue OF research was surly a motivating factor 6
position to the extreme secularism OF the Young Turk regime. Given the combi 6
rve to underline this significance OF unity. In contrast to an engraving mac 6
litics, determining the successful OF a civilian administration. In conclusi 6
However the degree OF superiority OF the former over the latter in terms OF 6
ct!" Ranke criticizes the validity OF some OF the speeches that Guicciardini 6
p forward and subsequent low level OF technology, Sino-Soviet split and resu 6
there is an ever expanding circle OF women in the capital cities OF Serbia, 7
ated nor retorted. In 1988 a crowd OF enraged Congolese women with photos OF 7
"Islam" brings a greater backlash OF Islamism. Bernard Lewis, The Crisis OF 7
deed not, because it's a lifestyle OF elites. Are we to understand that wher 7
undergoing a humiliating ceremony OF re-acceptance into the church, joined 7
der and guide OF society, a period OF violent "rocking", shocks, struggle ar 7
s would "develop into repositories OF knowledge to ensure the continuation C 7
y. The New diplomacy is an example OF cross-border relations between state a 7
external threats; a recent example OF this being EU sanctions on imported cl 7
theory by exploring how relations OF dependency in key global economic stru 7
rable. 'The manifold relationships OF force that take shape and come into pl 7
reation OF a European unified army OF 100000 soldiers, a proposal that woul 7
had been chosen from all chemists OF all ages to discover natural transmuta 7
aphic location, rather as a symbol OF a way OF life or values which threaten 7
ture OF the state prompts a system OF inclusion and exclusion, which separat 7
r and frames Foucault's conception OF resistance in the same way - as a two- 8
ges the foundations OF the concept OF just war itself. It weakens the self-i 8
mber 11th environment, the concept OF anticipatory self-defence requires no 8
y built on the prominent mythology OF martyrdom in Irish political history. 8
ovement sought to redefine notions OF identities related to the nation, ethr 8
OF exclusionary politics to notions OF differences, with his theory on the 'p 8
bate, it is the Rawlsian principle OF justice as fairness which does most to 8
Hutu ethnicity overrode any sense OF sisterhood with Tutsi women hence Hutu 8
alism. More specifically, theories OF original accumulation attempt to pin-p 8
n the flipside, Rawls's conception OF "justice as fairness" has two fundamer 8
looks at the religious conceptions OF women that had endured for centuries, 8
OF no-one. With a new understanding OF power we truly perceive the influence 8
ll upon the innocent as every form OF punishment meted out by the criminal 9

| | | |
|-------------------------------------|--|----|
| becomes the most important form OF | domination in global politics. Consequ | 9 |
| during this time that the style OF | painting known as Neoclassicism came t | 9 |
| nd invest into non-violent means OF | mediation and integration . Even thoug | 9 |
| ureaucratic and deficient system OF | identifying a conflict as genocide, ac | 9 |
| ese tribunals changed the system OF | accountability in conflict situations. | 9 |
| irming this interplay and method OF | evaluation. Jungk's book had provoked | 9 |
| r. This extremely precise method OF | research with its need to verify the b | 9 |
| s an extremely scientific method OF | writing history - avoiding the temptat | 9 |
| ting and enforcing optimal modes OF | individual conduct' <(Coleman & Tucker | 9 |
| nding OF the East. It was 'a way OF | coming to terms with the Orient that i | 9 |
| is often regarded as the father OF | economics, and his writings have been | 10 |
| y to anoint Ranke as the 'father OF | scientific history'. This was due to h | 10 |
| guided to call Ranke the 'father OF | scientific history'. Vierhaus, 'Histor | 10 |
| himself as the global figurehead OF | the communist movement. Mao did not be | 10 |
| de Montfort's election as leader OF | the crusade. The count was a complex c | 10 |
| al, economic, and social leaders OF | the Czechs would be held ultimately re | 10 |
| OF the previous Byzantine rulers OF | Semitic lands. Such a contrast cannot | 10 |
| n and the occupation OF the head OF | the household. The term servant was us | 10 |
| after the Fourth Lateran Council OF | 1216, in which de Monfort was given th | 11 |
| relationship with it. The events OF | 1789-92 brought meaning to his Horatii | 11 |
| to depict the Tennis Court Oath OF | 1789. Dubois-CrancÃ© desired David as | 11 |
| ithout the Bad Godesberg reforms OF | 1959 - this was a conference where the | 11 |
| iod leading up to the Revolution OF | 1789. It was during this time that the | 11 |
| anian Revolution, Iran-Iraq war OF | 1980-88, Gulf war OF 1990-91 are suffi | 11 |
| eavily influenced by the effects OF | enlargement and there are two main the | 12 |
| and dominating power. The effect OF | this however is to limit what can and | 12 |
| the UAR, and the limited impact OF | his socialist policies in bringing gre | 12 |
| White argues that the influence OF | the railway on London's 'hypergrowth' | 12 |
| state. With such deep influences OF | religious conviction and involvement w | 12 |
| ew insight for feminist accounts OF | women's agency in conflicts. By castir | 13 |
| that history must be the history OF | the inter-action between Mastery and S | 13 |
| ese paintings and David's sketch OF | Marie Antoinette on the way to the gui | 13 |
| a Canon OF medicine - a summary OF | Galenic medicine practice that is thou | 13 |
| ow-how, amongst many other types OF | knowledge. From each body OF knowledge | 14 |
| e radicalism came from two types OF | worker; 'advanced' and 'backward'. 'Ac | 14 |
| me to establish the penal colony OF | Botany Bay. And the extent to which it | 15 |
| e Ministry OF Commerce, Ministry OF | Defense elite as well as the main oil | 15 |
| eginning OF the Federal Republic OF | Germany, more specifically in the 1950 | 15 |
| ne Bay OF Biscay and the Straits OF | Gibraltar, meaning that a friendly Fra | 15 |
| nnment and the Liberation Tigers OF | Tamil Eelam (LTTE), a secessionist gro | 15 |
| cteristics OF tomorrow's balance OF | power between China and the US. Under | 15 |
| nerently within the jurisdiction OF | the national state. This point has bee | 16 |
| e workers wanted pay for the day OF | Nicholas II's coronation in May 1896 a | 16 |
| n be seen to have drawn the gaze OF | the policymakers in Westminster to a n | 16 |
| stly over exaggerated the number OF | men that the Levellers would have excl | 16 |
| ith the queen would be something OF | an exaggeration. In particular, the qu | 16 |
| alogy OF morality in a condition OF | modern nihilism and exploring identity | 16 |
| heimer but it also gives a sense OF | moral weight and the paths OF duty tha | 16 |
| t in medieval times, yet because OF | women's legal invisibility they rarely | 16 |
| n it is fair to say that because OF | its complex institutional structure it | 16 |

| | |
|---|----|
| lude "the sending by or on behalf OF a state OF armed bands, groups, irregu | 16 |
| is used <u>for the personal benefit</u> OF the military fraternity, especially th | 16 |
| particularly noticeable in the case OF the daughter OF Al- Husain Ibn Abdawai | 16 |
| ariat rocketed out OF the control OF the nascent political organisations th | 16 |
| 'Nicodemism-recanting in the face OF overwhelming power" and warns us again | 16 |
| e, you will put it into the hands OF men to choose, [not] OF men [desirous] | 16 |
| ite being very much at the heart OF US military cold war defence programme | 16 |
| xperiences in conflicts. In midst OF their victimhood, women are often push | 16 |
| emocratic Republic. At the moment OF implementation Ostpolitik has created | 16 |
| olvement in Vietnam to the point OF full military commitment. Schulzinger | 16 |
| ebook.com/ OccupyWallSt in support OF the given definition. Here it should b | 16 |
| is with demonology. From the time OF England's conversion to Christianity i | 16 |
| 98: 129) It performs the function OF a mean OF communication and government | 16 |
| a result, 'it laid the foundation OF a world OF specialised labour, which t | 16 |
| ther relevant factor arising out OF leadership characteristics as well as | 16 |
| ld allow someone to push them out OF the way on the basis that you would wa | 16 |
| 'The Sultan was put in possession OF Jerusalem on Friday, the 27th OF Rejet | 16 |
| makes us ask important questions OF morality, action and intent that need | 16 |
| OF negative liberty he also talks OF economic slavery, for example, if a ma | 16 |
| that he "could not help thinking OF the fate OF Sodom: as indeed [he] did | 16 |
| ughters can we slip into thinking OF an external God." It not only seems ab | 16 |
| was in ruins that it was thought OF as fit to be transformed into a madhou | 16 |
| is not frightening; we are afraid OF what people will think OF and use it f | 16 |
| latter to succeed it was ignorant OF necessity to deny that 'without the gu | 16 |
| c forward, in favour OF the point OF view that leisure is more important th | 16 |
| itos. From an argumentation point OF view, that seems just inconsistent. To | 16 |
| d Germany under the Soviet sphere OF influence. In Germany, the imposition | 16 |
| although the system OF values was OF great importance to the Mediterranean | 16 |
| ribe the Albigensian crusade, and OF these three all naturally have their c | 16 |
| OF a calculable legal system and OF administration in terms OF formal rule | 16 |
| c only affects the disadvantaged. OF approximately 650 persons on death row | 16 |
| neral' can be seen as empiricist. OF this latter quality Ranke announced th | 16 |
| e current transition period it is OF outmost importance for the political f | 16 |
| / discretely. Avitus OF Vienne is OF primary importance in developing our u | 16 |
| sful and this was largely because OF de Monfort's experience and skill as a | 16 |
| ies OF the Enlightenment'. He is OF course referring to the method OF trans | 16 |
| ere, however, largely as a result OF having common enemies: Mongol, Western | 16 |
| anchise reform proposals in terms OF political reality and not abstract the | 16 |
| ith the USSR within the framework OF the 4-power agreement on Germany. 1947 | 16 |
| ssary section. Lewis, The Crisis OF Islam: Holy War and Unholy Terror, pp. | 17 |
| ockburn 'The Nature and Incidence OF Crime in England, 1559-1625: A Prelimi | 17 |
| Bolsheviks in Power: The Founding OF a New Order (2004/2005) pp.76-78 Read, | 17 |
| 70-1914', Bulletin OF the History OF Medicine, 79 (2005) p. 697 Micale, App | 17 |
| argues that The Cambridge History OF Islam is largely a 'chronology OF batt | 17 |
| ghteenth-Century France', Journal OF Modern History, Vol. 73 (March, 2001), | 17 |
| cients and neurasthenia', Journal OF the History OF Medicine, 32 (1977) p. | 17 |
| The Enlightenment and the Origins OF European Australia (Cambridge, 2002), | 17 |
| , p. 814. Tina Sideris, "Problems OF Identity, Solidarity and Reconciliatio | 17 |
| imilien Ruben, 'The Relationship OF Bolshevism to Marxism', in Pipes, Rev | 17 |
| MacPherson, The Political Theory OF Possessive Individualism, pp. 122 - 12 | 17 |